



Pettigrew, R. (2018). The Population Ethics of Belief: In Search of an Epistemic Theory X. *Noûs*, 52(2), 336-372.  
<https://doi.org/10.1111/nous.12164>

Peer reviewed version

License (if available):  
CC BY-NC

Link to published version (if available):  
[10.1111/nous.12164](https://doi.org/10.1111/nous.12164)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Wiley at <http://onlinelibrary.wiley.com/doi/10.1111/nous.12164/abstract>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# The population ethics of belief: in search of an epistemic Theory X\*

Richard Pettigrew, University of Bristol<sup>†</sup>

## Abstract

Consider Phoebe and Daphne. Phoebe has credences in 1 million propositions. Daphne, on the other hand, has credences in all of these propositions, but she's also got credences in 999 million other propositions. Phoebe's credences are all very accurate. Each of Daphne's credences, in contrast, are not very accurate at all; each is a little more accurate than it is inaccurate, but not by much. Whose doxastic state is better, Phoebe's or Daphne's?

It is clear that this question is analogous to a question that has exercised ethicists over the past thirty years. How do we weigh a population consisting of some number of exceptionally happy and satisfied individuals against another population consisting of a much greater number of people whose lives are only just worth living? This is the question that occasions population ethics. In this paper, I go in search of the correct population ethics for credal states.

We start with an example:

**Birders** Consider two birders, Phoebe and Daphne. Phoebe has credences in 1 million propositions. All of them concern songbirds: they concern the habitats of particular songbirds, their range, varieties, markings, etc. Daphne, on the other hand, has credences in all of these propositions, but she's also got credences in 999 million other propositions. They concern other bird species in the passerine family besides the songbirds. They concern the habitats of these birds, their range, varieties, markings, etc. Phoebe's credences in all 1 million propositions to which she assigns credences are all very accurate in the following sense: her credences in true propositions are extremely high — she's nearly 100% confident in each — whereas her credences in false propositions are very low — she's nearly 0% confident in each. Each of Daphne's credences, in contrast, are not very accurate at all; each is a little more accurate than it is inaccurate, but not by much — she's between 50% and 51% confident in each truth and between 49% and 50% confident in each falsehood.

---

\***Acknowledgements:** A number of different people asked the questions that prompted me to write this paper; and the subsequent discussions helped enormously: in particular, Chris Nicholson, Zoe Johnson King, Keshav Singh, and Lindsay Brainard. I received a number of very helpful comments on earlier drafts from Daniel Greco, Jimmy Goodrich, Alejandro Pérez Carballo, Brian Talbot, Clayton Littlejohn, Samir Okasha, Ted Shear, and Ben Levinstein. I am also extremely grateful to an anonymous referee for this journal for their careful, thoughtful and extremely helpful suggestions. This work was supported by the European Research Council Seventh Framework Program (FP7/2007–2013), ERC grant 308961-EUT.

<sup>†</sup>Department of Philosophy, University of Bristol, Cotham House, Bristol, UK. BS6 6JL.

Whose doxastic state is better from the purely epistemic point of view, Phoebe's or Daphne's? That is, which has greater purely epistemic value?

Let us suppose, as we will throughout this paper, that the epistemic utility of a credence is given by its accuracy. A credence in a true proposition has greater accuracy — and thus, on this proposal, greater epistemic utility — the higher it is. And a credence in a false proposition has greater accuracy — and thus again, on this proposal, greater epistemic utility — the lower it is. That is, we will assume a sort of *veritism* about the epistemic value of credences (Joyce, 1998; Goldman, 2002). Of course, this is not the only account of epistemic utility available: we might say that the epistemic utility of a given credence in a proposition for a particular agent is the proximity of that credence to whichever credence is best supported by the agent's evidence — its proximity to the evidential probability of the proposition, if you like. Nonetheless, veritism is the account we will assume here. While I do think it is the correct account, my purpose in selecting it is not for its veracity; it is rather that this account is worked out in sufficient detail to make a useful case study of the sort of problem I wish to address here.

Now, let us return to Phoebe and Daphne. Daphne has opinions about vastly more propositions than Phoebe. But, by the lights of veritism, which we assume throughout, each of her credences is significantly worse than each of Phoebe's, epistemically speaking. Our question is just this: how do we weigh having credences with very high positive epistemic utility in some number of propositions against having credences with only slightly positive epistemic utility in some much greater number of propositions?

Now, it is clear that this question is analogous to a question that has exercised ethicists — and, in particular, utilitarians — over the past thirty years. How do we weigh a population consisting of some number of exceptionally happy and satisfied individuals against another population consisting of a much greater number of people whose lives are only just worth living? This is the question that occasions *population ethics*. As Derek Parfit noted, total utilitarianism in ethics — the view that the goodness of a population is simply the sum of the utilities of the individuals that comprise it — will sometimes say that the latter population is better than the former. He called that result the *repugnant conclusion* and took it to show that total utilitarianism isn't a viable theory in ethics. He then noted that certain natural alternatives also have unacceptable consequences. And he set ethicists the task of finding Theory X, a theory that gives a plausible ordering by their goodness of populations containing different individuals and different numbers of individuals (Parfit, 1984). In this paper, I go in search of Theory X for credal states. A putative epistemic Theory X will give an epistemic goodness ordering of credal states that are defined on different sets of propositions that come in different sizes.

I am not the first to embark on this search. Jennifer Carr (ta) has already furnished us with a host of valuable insights about the consequences of certain initially tempting orderings of credal states by their epistemic goodness. I begin in this paper by considering these insights and expanding on them. They appear to raise serious problems for Total Epistemic Utilitarianism, the epistemic Theory X that I will ultimately defend. However, before turning to that defence in Section 3.4, I wish to consider whether there is an alternative epistemic Theory X that escapes the objections that Carr raises. By appealing to analogues of impossibility theorems from population ethics, we can see that any epistemic Theory X must of necessity face substantial problems. I conclude by arguing that the problems with Total Epistemic Utilitarianism can be overcome.

## 1 Carr on epistemic utility theory

Before I consider Carr's observations in detail, I'd like to note a difference between my approach and hers. There are two ways to raise concerns about a putative ordering of certain items in terms of their goodness. On the first, we appeal directly to intuitions about the relative goodness of the items in question; we note that those intuitions conflict with the verdict given by the ordering we are criticizing; and we claim that this gives us reason to reject that ordering. For instance, take the case of population ethics. There, an example of this first way of objecting to a putative ordering of populations by goodness might run as follows: we note that the ordering ranks population *A* above population *B*; and we note that, intuitively, we take population *B* to be better than population *A*. This is the approach I will take to the population ethics of credences in Section 2. Carr, on the other hand, takes an alternative approach. She appeals to intuitions about what is rationally required or prohibited or permitted for someone who must choose between those items; she notes that those intuitions conflict with what we might think would follow about rationality if the ordering we are considering were the true ordering of those items by their goodness; and she worries that this gives us reason to reject that ordering. For instance, again take the case of population ethics. There, an example of this second style of objection to a putative ordering might run as follows: we note that the ordering ranks population *A* above population *B*; we infer that the ordering therefore tells us to choose to bring about *A* instead of *B* if those are the only two options available to us; and we note that, intuitively, it is at least rationally permissible to choose to bring about population *B* instead of population *A* given such a choice.

Thus, Carr and I both assume veritism, the account of epistemic utility on which the sole fundamental source of epistemic utility is accuracy. Carr adds to this theory of epistemic utility a theory of how facts about the rationality of credal states is determined by facts about their epistemic utility. In particular, she appeals to standard decision-theoretic principles to do that. This results in a species of epistemic consequentialism that has been called *epistemic decision theory* or *accuracy-first epistemology* (Joyce, 2009; Greaves, 2013; Pettigrew, 2016). We will have more to say about it below.

### 1.1 The framework of epistemic utility theory

In order to state Carr's concerns, let's introduce a little notation.

#### 1.1.1 Credence functions

We seek an ordering of credal states by their goodness. We represent a credal state by a *credence function*. An agent's credence function at a given time is a function  $c$  defined on a set of propositions  $\mathcal{F}$ . It takes each  $X$  in  $\mathcal{F}$  and returns  $c(X)$ , which gives the agent's credence in  $X$  at that time. By convention, we say that 0 is minimal credence and 1 is maximal credence. Thus, for each  $X$  in  $\mathcal{F}$ ,  $0 \leq c(X) \leq 1$ . We call  $\mathcal{F}$  the *opinion set* of  $c$ . We will assume throughout that  $\mathcal{F}$  is a finite set. But we will also assume that there are infinitely many propositions that lie outside  $\mathcal{F}$  and that might be added to it in order to expand it.

### 1.1.2 Global epistemic utility ordering

On the veritist or accuracy-first account of epistemic utility that we'll consider here, the epistemic utility of a credence function depends in part on the way the world is; in particular, it depends in part on whether the propositions to which the credence function assigns credences are true or false. Thus, what we seek is an ordering of credence functions for each way the world is. Thus, if  $w$  is a possible world, then we seek an ordering  $\preceq_w$  of credence functions by their epistemic goodness or utility at  $w$ . We write  $c \preceq_w c'$  if  $c'$  is at least as epistemically good as  $c$  at  $w$ . We will call  $\preceq_w$  the *global epistemic utility ordering* for  $w$ . And we will write  $c \prec_w c'$  if  $c \preceq_w c'$  but  $c' \not\preceq_w c$ ; and we will write  $c \sim_w c'$  if  $c \preceq_w c'$  and  $c' \preceq_w c$ .

### 1.1.3 Local epistemic utility function

In ethics, Theory X is an account of how the goodness of whole populations relates to the welfare or utility of the individuals that comprise them. This is true also in epistemology. Theory X is puzzling for those of us who hold that there is a sensible notion of purely epistemic utility, at least for individual doxastic states, such as particular credences assigned to particular propositions. Thus, the final component of our framework is a *local epistemic utility function*,  $s_X$ , for each proposition  $X$ . The function  $s_X$  takes a truth-value and a credence and returns the local epistemic utility of having that credence in  $X$  at a world at which it has that truth-value. By convention, we represent the truth-values numerically, so that 1 stands for truth and 0 for falsity. Thus, for  $0 \leq x \leq 1$ ,  $s_X(1, x)$  gives the local epistemic utility of having credence  $x$  in  $X$  when  $X$  is true, while  $s_X(0, x)$  gives the local epistemic utility of having credence  $x$  in  $X$  when  $X$  is false. A little notation that will be useful later: given a possible world  $w$  and a proposition  $X$ , we will write  $w(X)$  for the numerical truth-value of  $X$  at  $w$ , so that  $w(X) = 1$  if  $X$  is true at  $w$  and  $w(X) = 0$  if  $X$  is false at  $w$ . Thus, given a credence function  $c$  and a possible world  $w$ , the local epistemic utility at  $w$  of the credence that  $c$  assigns to  $X$  is  $s_X(w(X), c(X))$ .<sup>1</sup>

As we noted above, we are assuming veritism for credences in this paper. So it is reasonable to assume that the epistemic utility of a credence in a proposition at a world depends only on the credence, the proposition, and the truth-value of that proposition at the world in question. If we took a more evidentialist line, we'd need to allow the epistemic utility of a credence to vary with the agent's evidence as well. As well as assuming that an agent's epistemic utility is not a function of her evidence, the version of veritism we will consider here also assumes that the epistemic utility of an individual credence in a proposition is not a function of the process by which it was formed, nor a function of the other credences that the agent assigns. That is not to say that it denies that some processes are better than others; nor is it to say that the contribution of an individual credence to the overall epistemic utility of an agent's credal state does not depend on the other credences that comprise that credal state. Rather, this version of veritism assumes only that, whatever goodness a process has, it is determined by the epistemic utilities of the individual credences to which that process gives rise, rather than the other way round. And it assumes that the contribution made by an individual credence to the overall epistemic utility of a credal state is determined by taking the epistemic utility of that credence, as well as the epistemic utilities of all the other individual credences that comprise that state, and aggregating them appropriately. This leaves

<sup>1</sup>For the cognoscenti:  $s_X(i, x)$  can be thought of as the negative of a scoring rule (Predd et al., 2009).

open that the effect of each individual credence on the epistemic utility of the overall state might change depending on what other credences belong to that state.

Now, in what follows, we'll sometimes assume that our local epistemic utility function has certain properties. I list them here for ease of reference. The first property is *truth-directedness*.<sup>2</sup> This is a natural assumption if you think that the epistemic utility of a credence is determined only by its accuracy, as we do here. It says that the epistemic utility of a credence in a truth is an increasing function of that credence, while the epistemic utility of a credence in a falsehood is a decreasing function of that credence; and it says that maximal credence in a truth is better than maximal credence in a falsehood, while minimal credence in a falsehood is better than minimal credence in a truth.

**Truth-Directedness** For all propositions  $X$ ,

- (i)  $s_X(1, x)$  is a strictly increasing function of  $x$ ;
- (ii)  $s_X(0, x)$  is a strictly decreasing function of  $x$ ;
- (iii)  $s_X(0, 1) < 0 < s_X(1, 1)$  and  $s_X(1, 0) < 0 < s_X(0, 0)$

The second assumption we will make in some places below is that the local epistemic utility function is *strictly proper*. This means that, when we evaluate the expected epistemic utility of the various possible credences in a proposition  $X$  from the point of view of a particular given credence  $p$  in  $X$ , the given credence  $p$  expects itself to be epistemically the best out of all of them.

**Strict Propriety** For all propositions  $X$ , if  $0 \leq p \leq 1$ , then

$$ps_X(1, x) + (1 - p)s_X(0, x)$$

is maximized, as a function of  $x$ , at  $x = p$ .

This is a popular assumption in the literature on epistemic utility arguments for epistemic norms.<sup>3</sup> For instance, Predd et al. (2009) show that, together with Partial Additivity and Continuity (below), it can be used in the service of an argument for Probabilism, the norm that says that a rational agent will have a probabilistic credence function. We will have more to do with this argument below. Greaves & Wallace (2006) show that the other central Bayesian principle, Conditionalization, can also be derived using Strict Propriety together with some additional assumptions; Pettigrew (2014) shows that Strict Propriety combined with slightly different additional assumptions entails the Principle of Indifference; and Pettigrew (2013) motivates the Principal Principle from Strict Propriety together with further assumptions that are slightly different again. Joyce (2009, 279) sketches an argument in favour of a close relative of Strict Propriety that might be deployed from the point of view of many different accounts of epistemic utility, including more evidentialist ones; and Pettigrew (2016, Chapter 4) gives an alternative argument in favour of Strict Propriety from the point of view of veritism, in particular.

The third and final assumption we will make at various points below is that local epistemic utility is a continuous function of credence.

<sup>2</sup>The term is due to Joyce (2009), though he omits clause (iii) from the definition.

<sup>3</sup>Strict Propriety entails clauses (i) and (ii) of Truth-Directedness (Gneiting & Raftery, 2007, 363).

**Utility Continuity** For any proposition  $X$ ,  $s_X(0, x)$  and  $s_X(1, x)$  are both continuous functions of  $x$ .

Here are two families of local epistemic utility functions, each of whose members satisfy Truth-Directedness, Strict Propriety, and Utility Continuity. Their details won't concern us in what follows, but it might be useful to have them in mind. Both families are parameterized by two values,  $0 < \lambda \leq 1$  and  $0 \leq \alpha \leq 1$ .

- *Quadratic scoring rule*

(a)  $q_{\alpha, \lambda}(1, x) = \alpha - \lambda(1 - x)^2$ ;

(b)  $q_{\alpha, \lambda}(0, x) = \alpha - \lambda x^2$ .

- *Logarithmic scoring rule*<sup>4</sup>

(a)  $l_{\alpha, \lambda}(1, x) = \alpha + \lambda \ln x$ ;

(b)  $l_{\alpha, \lambda}(0, x) = \alpha + \lambda \ln(1 - x)$ .

## 1.2 Total Epistemic Utilitarianism

Having stated the framework in which we'll be working, we can turn to Carr's treatment of what we might call *total epistemic utilitarianism*:

**Total Epistemic Utilitarianism** If  $c$  and  $c'$  are credence functions defined on opinion sets  $\mathcal{F}$  and  $\mathcal{F}'$  respectively, then

$$c \preceq_w c' \text{ iff } \sum_{X \in \mathcal{F}} s_X(w(X), c(X)) \leq \sum_{X' \in \mathcal{F}'} s_{X'}(w(X'), c'(X'))$$

That is, one credence function is at least as good as another if the sum of the epistemic utilities of the credences assigned by the first is at least as great as the sum of the epistemic utilities of the credences assigned by the second.

Carr draws out unexpected and apparently unappealing consequences of this position. The theorem that follows (Theorem 2) is simply a generalization of the central observations in (Carr, *ta*, Sections 3.4-3.5). To state it, we require some terminology. First, the notion of an *epistemically neutral credence*.

**Definition 1** Suppose  $s$  is a local epistemic utility function,  $x$  is a credence, and  $X$  is a proposition. Then, if  $s_X(1, x) = s_X(0, x)$ , we say that  $x$  is an epistemically neutral credence in  $X$  relative to  $s$ .

That is, a credence is epistemically neutral if it has the same epistemic utility regardless of how the world turns out. For instance, for any member  $q_{\alpha, \lambda}$  of the family of quadratic scoring rules, 0.5 is the unique epistemically neutral credence, since  $q_{\alpha, \lambda}(1, 0.5) = q_{\alpha, \lambda}(0, 0.5)$ . And similarly for any member  $l_{\alpha, \lambda}$  of the family of logarithmic scoring rules. Now, we have the following result, which tells us that epistemically neutral credences always exist, providing we make certain assumptions about our local epistemic utility function.

---

<sup>4</sup> $\ln x$  is the natural logarithm of  $x$ . That is,  $\ln x := \log_e x$ . So  $\ln x$  is the power to which one must raise  $e$  to give  $x$  — that is,  $e^{\ln x} = x$  for  $x > 0$ .

**Proposition 1** *Suppose that  $\mathfrak{s}$  is truth-directed and continuous. Then every proposition has a unique epistemically neutral credence.*

(All proofs are given in the Appendix.) Given a proposition  $X$ , we denote its epistemically neutral credence  $r_X$ .

Our second piece of terminology. We say that a credence function  $c$  is *strictly dominated* if there is another credence function  $c^*$  such that  $c \prec_w c^*$  for all worlds  $w$ . That is,  $c$  is strictly dominated if there is  $c^*$  that is guaranteed to be better than  $c$ .<sup>5</sup>

Our third piece of terminology. We say that a proposition  $X$  is an *open proposition* if, for any probability  $0 \leq r \leq 1$ , there is some evidential situation in which an agent might find herself such that the unique rational response to that situation is to assign credence  $r$  to  $X$ . Thus, for instance, take the proposition *The coin will land heads on its next toss*. For any  $0 \leq r \leq 1$ , I might learn that the objective chance of that proposition is  $r$ , since I might learn that that is the bias of the coin towards heads. Thus, by the Principal Principle, in such a situation, I would be rationally required to assign credence  $r$  to that proposition (Mellor, 1971; Lewis, 1980; Joyce, 2009).

We can now state our first theorem.

**Theorem 2** *Suppose that  $\mathfrak{s}$  is truth-directed, strictly proper, and continuous. And suppose that the epistemic utility ordering  $\preceq_w$  is governed by Total Epistemic Utilitarianism. Then one of the following is true:*

- (a) *For all  $\varepsilon > 0$ , there is a proposition  $X$  whose epistemically neutral credence is below  $\varepsilon$  — that is,  $0 \leq r_X < \varepsilon$ ;*

OR

- (b) *There is a credence  $r$  and an open proposition  $X$  such that every credence function that assigns  $r$  to  $X$  is strictly dominated — that is, if  $c(X) = r$ , then there is  $c^*$  such that  $c \prec_w c^*$  for all worlds  $w$ .*

The proof — given in full detail in the Appendix — runs as follows. Either (i) there is an open proposition  $X$  whose epistemically neutral credence  $r_X$  is guaranteed to have negative epistemic utility; or (ii) every open proposition has an epistemically neutral credence that is guaranteed to have positive epistemic utility. We then show that, if (i), then (b) holds; and if (ii) holds and (a) is false, then (b) holds. If (i) holds, then any credence function that assigns  $r_X$  to  $X$  is strictly dominated by the credence function that results from removing  $X$  from its opinion set. If (ii) holds and (a) is false, then, for any credence function, it is always possible to choose a series of open propositions so that, if we add them all to the opinion set of that credence function and assign them their epistemically neutral credence, the resulting expanded credence function strictly dominates the original one. Since this holds of any credence function, there is certainly a credence  $r$  and an open proposition  $X$  such that every credence function that assigns  $r$  to  $X$  is strictly dominated by an expanded credence function, and thus strictly dominated, as required by (b).

<sup>5</sup>To avoid confusion, note that there is a possible alternative use of this terminology that is analogous to certain uses in the population ethics literature. We might say that, if  $c^*$  and  $c$  are credence functions defined on the same set of propositions, then  $c^*$  strictly dominates  $c$  at a particular world  $w$  if, for each proposition to which  $c$  and  $c^*$  assign a credence, the epistemic utility at that world of the credence that  $c^*$  assigns to it is higher than the epistemic utility at that world of the credence that  $c$  assigns to it. This is not our usage, and the terminology will not be employed in this way in our article.



Thus, if we make the assumptions on local epistemic utility functions that are standardly made in the literature on accuracy-first or epistemic utility arguments for epistemic norms, and if we assume that one credence function is epistemically at least as good as another if the total epistemic utility of the first is at least as great as the total epistemic utility of the second, then we are faced with a dilemma between two unappealing consequences.

- First possibility (a). According to (a), there are propositions whose epistemically neutral credence is as close to 0 as we wish. This is bizarre and certainly flies in the face of veritism. It means, for instance, that there is a proposition such that a credence of 0.0000001 in that proposition is just as epistemically good if the proposition is true as if it's false. Another consequence: There is a proposition and a credence in that proposition such that the credence would sanction taking a bet that wins you £1 if the proposition is false and loses you £1 billion if the proposition is true, but which is equally good epistemically speaking whether or not the proposition is true or false.
- Second possibility (b). According to (b), there is a credence and an open proposition such that, if the credence is assigned to the proposition, then the resulting credence function is strictly dominated. This, too, is bizarre. After all, in decision theory, if one option is dominated by another, the first is taken to be irrational. Thus, according to this second possibility, there is a credence and an open proposition such that it is always irrational to assign that credence to that proposition. But, by the definition of an open proposition, there are situations in which an agent is not only rationally permitted to assign that credence to that proposition but is rationally obliged to.

This second possibility (b) seems particularly troubling for those who would appeal to epistemic utility to justify principles such as Probabilism. After all, the accuracy-first argument for Probabilism given by Joyce (1998, 2009) is based on the following result:

- (I) For every non-probabilistic credence function  $c$ , there is a probabilistic credence function  $c^*$  *defined on the same opinion set as  $c$*  that is guaranteed to be better than  $c$ .
- (II) For every probabilistic credence function  $c$ , there is no credence function  $c^*$  *defined on the same opinion set as  $c$*  that is guaranteed to be at least as good as  $c$ .

He then hopes to apply a decision-theoretic principle to derive Probabilism. However, to apply such a principle, he would need something stronger than (I) and (II). He would need:

- (I') For every non-probabilistic credence function  $c$ , there is a probabilistic credence function  $c^*$  that is guaranteed to be better than  $c$ .
- (II') For every probabilistic credence function  $c$ , there is no credence function  $c^*$  that is guaranteed to be at least as good as  $c$ .

(I') follows from (I), but the second possibility (b) above directly contradicts (II'). It tells us that there is a credence  $r$  and a proposition  $X$  such that every credence function that assigns  $r$  to  $X$  is strictly dominated. But of course there will be probability functions amongst those credence functions that assign  $r$  to  $X$ . So there will be probability functions that are strictly dominated, *contra* (II'). I will argue in Section 3.4 that this does not in fact cause a problem for Joyce's argument. Indeed, my main concern with Carr's approach is that I do not think that Total and Average Epistemic Utilitarianism have

the consequences for rationality that are suggested by Theorem 2 above and Theorem 3 below.

### 1.3 Average Epistemic Utilitarianism

A natural alternative to Total Epistemic Utilitarianism is Average Epistemic Utilitarianism: whereas the former orders credence functions by their total epistemic utility, the latter orders them by their average epistemic utility.

**Average Epistemic Utilitarianism** If  $c$  and  $c'$  are credence functions defined on  $\mathcal{F}$  and  $\mathcal{F}'$  respectively, then

$$c \preceq_w c' \text{ iff } \frac{1}{|\mathcal{F}|} \sum_{X \in \mathcal{F}} s_X(w(X), c(X)) \leq \frac{1}{|\mathcal{F}'|} \sum_{X' \in \mathcal{F}'} s_{X'}(w(X'), c'(X'))$$

where  $|\mathcal{F}|$  is the number of propositions in  $\mathcal{F}$ , and  $|\mathcal{F}'|$  is the number of propositions in  $\mathcal{F}'$ .<sup>6</sup>

Average Epistemic Utilitarianism avoids the problems raised by Theorem 2 above.<sup>7</sup> However, Average Epistemic Utilitarianism suffers from a problem similar to the one raised for Total Epistemic Utilitarianism by Theorem 2(b). Suppose that having maximal credence in a truth is at least as good as having minimal credence in a falsehood. Then, as Carr points out, if I have credences only in tautologies, and if I assign credence 1 to each of these tautologies, then Average Epistemic Utilitarianism seems to prohibit me from assigning non-extremal credence to any new proposition. After all, such a move is guaranteed to reduce my average accuracy. But, as we noted above, for any credence and any proposition, there is a situation in which it is rationally permitted to have that credence in that proposition. And, mutatis mutandis, if having minimal credence in a falsehood is at least as good as having maximal credence in a truth. Another way to put this is to note the following theorem:

**Theorem 3** *Suppose that  $s$  is truth-directed. And suppose that the epistemic utility ordering  $\preceq_w$  is governed by Average Epistemic Utilitarianism. Then, any credence function that assigns any non-extremal credences is strictly dominated.*

If we apply the dominance principle that we introduced above, it follows from Theorem 3 that, in the presence of Average Epistemic Utilitarianism, every non-extremal credence function — that is, every credence function that assigns at least some non-extremal credences — is irrational. This is analogous to the usual worry about the version of utilitarianism in ethics that is analogous to Average Epistemic Utilitarianism: that position entails that, if I have a population of a million deliriously happy people, it is impermissible to create a new person who I know will only be very, but not deliriously, happy.

As with Total Epistemic Utilitarianism, Average Epistemic Utilitarianism seems to have unacceptable consequences for rationality. In fact, as I mentioned above, I will deny that they really have these consequences. I will instead make trouble for these proposals and many

---

<sup>6</sup>Recall: we are restricting attention to credence functions with finite opinion sets in this article. Thus,  $\frac{1}{|\mathcal{F}|}$  and  $\frac{1}{|\mathcal{F}'|}$  are always well-defined.

<sup>7</sup>The reason is this: if you look to the proof of Theorem 2 in the Appendix below, equation (2) does not entail that  $c^*$  strictly dominates  $c$  when we assume Average Epistemic Utilitarianism instead of Total Epistemic Utilitarianism.

others by identifying unintuitive features of the epistemic utility orderings themselves, independent of their consequences for rationality. Indeed, I will show that trouble looms for *any* putative epistemic Theory X, just as happens in the ethical case. I will end by arguing that Total Epistemic Utilitarianism is nonetheless the best population ethics for credences.

## 2 The Paradox of Epistemic Utility for Credences

In this section, we are concerned not with assessing putative epistemic utility orderings by looking to their consequences concerning what is rational; instead, we will assess them directly by looking at how they order certain pairs of credence functions. We are thus assessing them in the way that orderings of populations are often assessed in the most well-known arguments in population ethics. Indeed, we'll be considering an adaptation of the epistemic analogue of Parfit's Mere Addition Paradox. Our paradox — which we will call the Paradox of Epistemic Utility for Credences — consists of four plausible but jointly inconsistent principles that are taken to govern our epistemic utility ordering of credence functions.

### 2.1 Partial Additivity

The first of our four plausible principles that together constitute this Paradox states that, for a particular class of cases in which Total and Average Epistemic Utilitarianism are guaranteed to agree, they are right. These are the cases in which the two credence functions that we are comparing are defined on the same opinion set:

**Partial Additivity** If  $c, c'$  are credence functions defined on  $\mathcal{F}$ , then

$$c \preceq_w c' \text{ iff } \sum_{X \in \mathcal{F}} s_X(w(X), c(X)) \leq \sum_{X \in \mathcal{F}} s_X(w(X), c'(X))$$

Thus, if Debby and Lili both have credences in the same propositions, then Debby's credence function is at least as good as Lili's just in case her total epistemic utility is at least as high as Lili's (and, equivalently, her average epistemic utility is at least as high as Lili's).

Note that the ethical analogue of this principle is subject to the usual objections to utilitarianism in ethics, namely, that it permits us to trade-off the misery of the few in favour of the joy of the many. As a result, many population ethicists would not assume a principle like this. For them, equality is an important virtue of a population and Partial Additivity pays no attention to this. Rather, they assume something like **Non-Anti-Egalitarianism**, which says that, if  $A$  and  $B$  are two populations of the same size and if  $A$  has at least as great total and average utility as  $B$  *whilst also being at least as equal as  $B$* , then  $A$  is at least as good as  $B$ , all things considered (Ng, 1989; Arrhenius, 2000; Huemer, 2008).<sup>8</sup> We need have no such qualms and we need not qualify Partial Additivity in this way. There is no extrinsic or intrinsic value to having credences with equal epistemic utility. And indeed, at least in the presence of veritism, it would be a bizarre claim. After all, suppose I have credences concerning the outcome of a fair lottery with 1 million tickets. Then, if  $w_i$  is the proposition that ticket  $i$  wins, we presumably think that it would be at least as good for me to assign a credence of  $\frac{1}{1,000,000}$  to each  $w_i$ ,  $\frac{2}{1,000,000}$  to each  $w_i \vee w_j$  (for distinct  $i, j$ ),  $\frac{3}{1,000,000}$  to each  $w_i \vee w_j \vee w_k$  (for

<sup>8</sup>The name of the principle is due to Ng (1989, 238) and adopted by Arrhenius (2000, 253). In discussion of distributive justice, this principle is sometimes known as the *Pigou-Dalton Principle* (Pigou, 1912; Dalton, 1920).

distinct  $i, j, k$ ), and so on, as to assign a credence of  $\frac{1}{2}$  to each  $w_i$ ,  $\frac{1}{2}$  to each  $w_i \vee w_j$ ,  $\frac{1}{2}$  to each  $w_i \vee w_j \vee w_k$ , and so on. However, if we favour equality of credences, the latter will often be preferable. For many epistemic utility functions, all of the credences in the latter assignment will have the same epistemic utility, since, according to those epistemic utility functions, a credence of  $\frac{1}{2}$  in any truth has exactly the same epistemic utility as a credence of  $\frac{1}{2}$  in any falsehood — that is, in the language of Section 1.2,  $\frac{1}{2}$  is the epistemically neutral credence for every proposition. On the other hand, for any truth-directed epistemic utility function, the credences in the former assignment will have epistemic utilities that span nearly the whole gamut of possible epistemic utility values from extremely accurate to extremely inaccurate, since it will include a credence of  $\frac{1}{1,000,000}$  in a truth and the same credence in a falsehood, and it will include a credence of  $\frac{999,999}{1,000,000}$  in a falsehood and the same credence in a truth. Thus, unlike in the ethical case, it seems that equality is no virtue in the credal case.

A major virtue of Partial Additivity is its role in epistemic arguments for various principles of credal rationality, such as Probabilism, Conditionalization, the Principal Principle, and so on. The strongest argument for each of these principles assumes Strict Propriety, Utility Continuity, and either Partial Additivity or a strengthening of it that concerns the ordering of credence functions by their expected epistemic utility relative to some probability distribution (Predd et al., 2009; Greaves & Wallace, 2006; Pettigrew, 2013, 2014).<sup>9</sup> While this doesn't constitute a direct argument in favour of Partial Additivity, it does illustrate the cost of denying it.

## 2.2 Transitivity

One consequence of Partial Additivity is that  $\preceq_w$  is a transitive ordering on every set of credence functions all of whose members are defined on the same opinion set: that is, if  $c, c', c''$  are defined on the same opinion set, and if  $c \preceq_w c'$  and  $c' \preceq_w c''$ , then  $c \preceq_w c''$ . Our next principle extends this latter claim to the whole universe of possible credence functions:

**Transitivity** Suppose  $c, c'$ , and  $c''$  are credence functions. Then, if  $c \preceq_w c'$  and  $c' \preceq_w c''$ , then  $c \preceq_w c''$ .

If a relation orders items by their all-things-considered goodness, then it is often claimed that the transitivity of that relation is an analytic truth (Broome, 2004, Section 4.1). In Section 3.3, we'll consider whether we might escape the Paradox of Epistemic Utility for Credences by denying transitivity in the way that Larry Temkin (1987) tries to escape the Mere Addition Paradox by denying transitivity in the ethical case. At that point, we'll take issue with the claim that Transitivity is an analytic truth either in the epistemic or the ethical case.

## 2.3 No Repugnance

Our third principle denies the epistemic analogue of what is called in ethics the *very repugnant conclusion* (Arrhenius, 2003). This a consequence of total utilitarianism that is apparently

<sup>9</sup>The required strengthening is likely to look like this, where we are now ordering credence functions not relative to a possible world, but rather relative to a probability distribution over all the possible worlds:

**Partial Additivity\*** If  $c, c'$  are credence functions defined on  $\mathcal{F}$ , and  $p$  is a probability function, then

$$c \preceq_p c' \text{ iff } \sum_w p(w) \sum_{X \in \mathcal{F}} \mathfrak{s}_X(w(X), c(X)) \leq \sum_w p(w) \sum_{X \in \mathcal{F}} \mathfrak{s}_X(w(X), c'(X))$$

even more repugnant than the Repugnant Conclusion itself. In ethics, the Repugnant Conclusion says that, for any degrees of utility  $0 < s < h$ , and for any population all of whose members have utility higher than  $h$ , there is a better population all of whose members have utility lower than  $s$ . The conclusion becomes more and more repugnant, it seems, as we set  $h$  higher and higher and  $s$  lower and lower (but still positive). Thus, one consequence of the Repugnant Conclusion is that, for any population of deliriously happy people, there is a better population all of whose members live lives that are only just worth living. The Very Repugnant Conclusion goes further: it says that, for any degrees of utility  $l < 0 < s < h$ , and any population all of whose members have utility higher than  $h$ , and any finite positive number  $n$ , there is a better population consisting of  $n$  individuals with utility lower than  $l$  and whose remaining individuals have utility lower than  $s$ . Again, we make this more repugnant by setting  $h$  and  $n$  higher, and  $s$  and  $l$  lower. One consequence of the Very Repugnant Conclusion is that, for any population of deliriously happy people, and any population of people enduring appalling suffering, the latter population can be made to be better than the former just by adding people whose lives are only just worth living. The Repugnant and Very Repugnant Conclusions both follow from total utilitarianism in ethics; and their epistemic analogues follow from Total Epistemic Utilitarianism. Thus, Total Epistemic Utilitarianism entails that, for any degrees of epistemic utility  $l < 0 < s < h$ , and any credence function all of whose credences have epistemic utility higher than  $h$ , and any finite positive number  $n$ , there is a better credence function that assigns  $n$  credences that have epistemic utility lower than  $l$  and whose remaining credences have epistemic utility lower than  $s$ . This is the epistemic analogue of the Very Repugnant Conclusion. Here is an instance: Suppose Phoebe has 1 million credences, all of which are extremely accurate. Then there is a credence function that Daphne might have that assigns 1 trillion credences that are extremely inaccurate and whose remaining credences are only very slightly more accurate than inaccurate, such that Daphne's credence function is better than Phoebe's. This claim seems to conflict with our epistemic intuitions. The conflict may not be as strong as in the ethical case. In the ethical case, many people respond with revulsion to the claim that we can always compensate for the extreme suffering of an enormous number of people by producing an even more enormous number of people whose lives are only just worth living. The same response is not appropriate in the epistemic case. But nonetheless I think there is an intuitive rejection of the epistemic analogue of such a claim, such as that exemplified by the ranking of Daphne above Phoebe in the example just given. Thus, for the moment, we deny it and assume the following principle, though as I have mentioned already, we will ultimately accept Total Epistemic Utilitarianism and the epistemic version of the Very Repugnant Conclusion that follows from it. So, for those who do not share the intuition that the epistemic version of the Very Repugnant Conclusion is indeed repugnant, I will ultimately argue that your intuitive response gets things exactly right.

**No Repugnance** There are levels of epistemic utility  $l < 0 < s < h$ , a credence function  $c^+$  defined on  $\mathcal{F}^+$ , and a number  $n$ , such that the following holds:

- (i) each credence that  $c^+$  assigns has epistemic utility higher than  $h$ ;
- (ii) if  $c$  is a credence function defined on  $\mathcal{F} \supseteq \mathcal{F}^+$  and  $|\mathcal{F}| > n$ , and  $n$  of the credences that  $c$  assigns have epistemic utility below  $l$ , and the remaining credences have epistemic utility below  $s$ , then  $c^+ \succeq_w c$ .

Thus, again assuming veritism, No Repugnance says that there's a credence function that

assigns only very accurate credences that is at least as good as any credence function that assigns very inaccurate credences to  $n$  propositions and only slightly accurate credences to all the rest. Total Epistemic Utilitarianism entails Partial Additivity and Transitivity (above) and Benign Addition (below), but it is incompatible with No Repugnance.

## 2.4 Benign Addition

Our final principle is this:

**Benign Addition** Suppose  $c$  and  $c'$  are credence functions on  $\mathcal{F}$  and  $\mathcal{F}'$  respectively. And suppose  $\mathcal{F} \subseteq \mathcal{F}'$ . Then, if

- (i) For each proposition  $X$  that is in  $\mathcal{F}'$  and also in  $\mathcal{F}$ , the credence that  $c'$  assigns to  $X$  has higher epistemic utility at  $w$  than the credence that  $c$  assigns to  $X$ , and
- (ii) For each proposition  $X$  that is in  $\mathcal{F}'$  but not also in  $\mathcal{F}$ , the credence that  $c'$  assigns to  $X$  has positive epistemic utility at  $w$ ,

then  $c \preceq_w c'$ .

The idea is this. Suppose I change my credence function in two ways: first, I change each of my current credences in a way that makes all of my new credences in those propositions better than all of my old credences in those propositions; second, I assign credences to a bunch of new propositions in a way that makes all of those new credences have positive epistemic utility. Then, according to Benign Addition, my new credence function is at least as good as my old one. Average Epistemic Utilitarianism entails Partial Additivity, Transitivity, and No Repugnance (above), but it is incompatible with Benign Addition.

## 2.5 The Paradox of Epistemic Utility for Credences

Finally, we can state the Paradox of Epistemic Utility for Credences. It comes in the form of the following theorem:

**Theorem 4** *There is no epistemic utility ordering  $\preceq_w$  that satisfies Partial Additivity, Transitivity, No Repugnance, and Benign Addition.*

All other proofs in this paper are given in the Appendix. However, it will be useful to see how this proof works so that we can refer back to particular moves later on. For that reason, we'll include it here.

*Proof of Theorem 4.* We begin with the items posited by No Repugnance. These are:

- (i) a series of levels of epistemic utility  $l < 0 < s < h$ ;
- (ii) a number  $n$ ; and
- (iii) a credence function  $c^\dagger$  defined on a set of propositions  $\mathcal{F}^\dagger = \{X_1, \dots, X_q\}$  such that  $s_{X_i}(w(X_i), c^\dagger(X_i)) > h$  for all  $X_i$  in  $\mathcal{F}^\dagger$ .

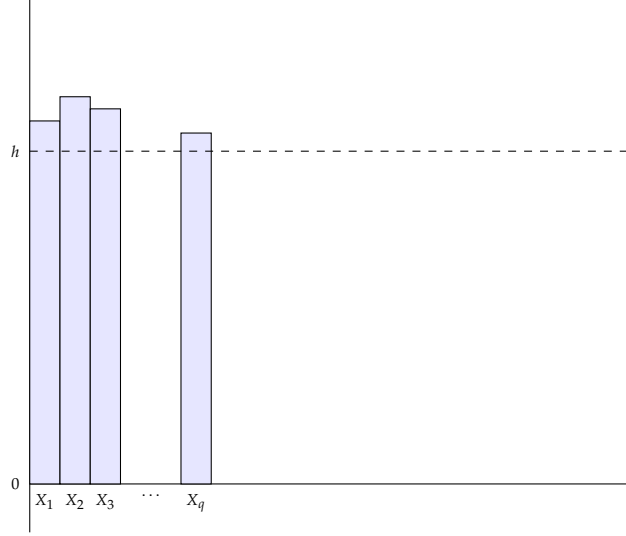


Figure 1: We plot the epistemic utility of the credences assigned by  $c^+$  to the propositions  $X_1, \dots, X_q$  in  $\mathcal{F}^+$ .

We illustrate  $c^+$  in Figure 1.

Now pick  $\varepsilon > 0$  and  $l' < l$  and  $0 < s'' < s' < s$ . And then choose  $m$  so that

$$\sum_{i=1}^q \mathfrak{s}_{X_i}(w(X_i), c^+(X_i)) + \varepsilon q + (m - q)s'' < nl' + (m - n)s' \quad (1)$$

By the Archimedean property of the reals, there is such an  $m$ .

Now define  $c^{++}$  on the set of propositions  $\mathcal{F}^{++} = \{X_1, \dots, X_q, \dots, X_n, \dots, X_m\}$  so that

- (i)  $\mathfrak{s}_{X_i}(w(X_i), c^{++}(X_i)) = \mathfrak{s}_{X_i}(w(X_i), c^+(X_i)) + \varepsilon$  for  $1 \leq i \leq q$ ; and
- (ii)  $\mathfrak{s}_{X_i}(w(X_i), c^{++}(X_i)) = s''$  for  $q < i \leq m$ .

We illustrate  $c^{++}$  in Figure 2.

Next, define  $c^*$  on  $\mathcal{F}^{++}$  so that:

- (i)  $\mathfrak{s}_{X_i}(w(X_i), c^*(X_i)) = l'$  for  $1 \leq i \leq n$ ; and
- (ii)  $\mathfrak{s}_{X_i}(w(X_i), c^*(X_i)) = s'$  for  $n < i \leq m$ .

We illustrate  $c^*$  in Figure 3.

Then, by (1),

$$\begin{aligned} \sum_{i=1}^m \mathfrak{s}_{X_i}(w(X_i), c^{++}(X_i)) &= \sum_{i=1}^q \mathfrak{s}_{X_i}(w(X_i), c^+(X_i)) + \varepsilon q + (m - q)s'' \\ &< nl' + (m - n)s' \\ &= \sum_{i=1}^m \mathfrak{s}_{X_i}(w(X_i), c^*(X_i)) \end{aligned}$$

So, by Partial Additivity,  $c^{++} \prec_w c^*$ , since both credence functions are defined on the same set of propositions. By Benign Addition,  $c^+ \preceq_w c^{++}$ . And so, by Transitivity,  $c^+ \prec_w c^*$ . And this latter inequality contradicts No Repugnance, which gives our contradiction.  $\square$

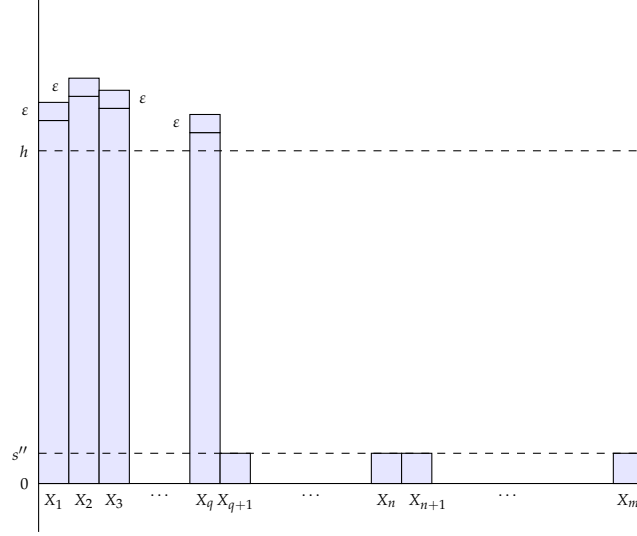


Figure 2: We plot the epistemic utility of the credences assigned by  $c^{++}$  to the propositions  $X_1, \dots, X_m$  in  $\mathcal{F}^{++}$ .

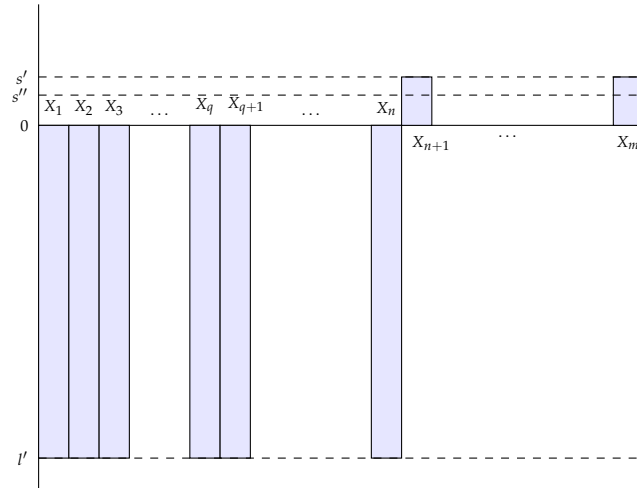


Figure 3: We plot the epistemic utility of the credences assigned by  $c^*$  to the propositions  $X_1, \dots, X_m$  in  $\mathcal{F}^{++}$ .



### 3 Responses to the paradox

In the next four sections, we will consider four possible responses to the Paradox stated in the previous section. We will consider whether we might deny Benign Addition, Partial Additivity, Transitivity, or No Repugnance.

#### 3.1 Denying Benign Addition

Let's begin with Benign Addition. And let us begin by noting that this principle is incompatible with Average Epistemic Utilitarianism. Benign Addition is also incompatible with the epistemic analogue of Ng's Variable Value Principle (Ng, 1989). To introduce this principle, we note that Total and Average Epistemic Utilitarianism are both members of the following family of principles, which are parameterized by a function  $f : \mathbb{N} \rightarrow \mathbb{R}^+$ :

**Epistemic Utilitarianism<sub>f</sub>** If  $c$  and  $c'$  are credence functions defined on  $\mathcal{F}$  and  $\mathcal{F}'$  respectively, then

$$c \preceq_w c' \text{ iff } \frac{f(|\mathcal{F}|)}{|\mathcal{F}|} \sum_{X \in \mathcal{F}} s_X(w(X), c(X)) \leq \frac{f(|\mathcal{F}'|)}{|\mathcal{F}'|} \sum_{X' \in \mathcal{F}'} s_{X'}(w(X'), c'(X'))$$

Total Epistemic Utilitarianism is Epistemic Utilitarianism<sub>f</sub> when  $f(n)$  is a non-constant linear function of  $n$ . Average Epistemic Utilitarianism is Epistemic Utilitarianism<sub>f</sub> when  $f(n)$  is a constant  $k$ . The epistemic analogue of Ng's Variable Value Principle (with parameter  $f$ ) is Epistemic Utilitarianism<sub>f</sub> when  $f$  is a bounded, strictly increasing, strictly concave function. The idea is that, when comparing two credence functions both defined on small opinion sets, the Variable Value Principle (with  $f$ ) demands that they are ordered much as Total Epistemic Utilitarianism would have them ordered; however, when comparing two credence functions both defined on large opinion sets, the principle demands that they are ordered pretty much in line with the demands of Average Epistemic Utilitarianism. Thus, the Variable Value Principle is supposed to secure the best of both earlier principles. However, like Average Epistemic Utilitarianism, it violates Benign Addition.

Perhaps this gives us reason to abandon Benign Addition? I think not. After all, it turns out that Average Epistemic Utilitarianism and any instance of the epistemic analogue of the Variable Value Principle also violate the following principle (Arrhenius, 2000, 251):

**No Sadism** There are epistemic utility levels  $l < 0 < s < h$  such that the following is true. Suppose  $c, c', c''$  are credence functions defined on  $\mathcal{F}, \mathcal{F}', \mathcal{F}''$  respectively, where  $\mathcal{F}$  and  $\mathcal{F}'$  are disjoint,  $\mathcal{F}$  and  $\mathcal{F}''$  are disjoint, and  $\mathcal{F}' \subseteq \mathcal{F}''$  or  $\mathcal{F}'' \subseteq \mathcal{F}'$ . Then, if

- (i) Each credence that  $c$  assigns has epistemic utility above  $h$  at  $w$ ,
- (ii) Each credence that  $c'$  assigns has positive epistemic utility below  $s$  at  $w$ ,
- (iii) Each credence that  $c''$  assigns has epistemic utility below  $l$  at  $w$ ,

then  $cc'' \preceq_w cc'$ .<sup>10</sup>

---

<sup>10</sup>If  $c, c'$  are defined on  $\mathcal{F}, \mathcal{F}'$ , respectively, and  $\mathcal{F}, \mathcal{F}'$  are disjoint, then we define a new credence function  $cc'$  on  $\mathcal{F} \cup \mathcal{F}'$  as follows:  $cc'(X) = c(X)$  if  $X$  is in  $\mathcal{F}$  and  $cc'(X) = c'(X)$  if  $X$  is in  $\mathcal{F}'$ .

That is, if, for all  $X$  in  $\mathcal{F}$ ,  $X'$  in  $\mathcal{F}'$ , and  $X''$  in  $\mathcal{F}''$ ,

$$\mathfrak{s}_{X''}(w(X''), c(X'')) < l < 0 < \mathfrak{s}_{X'}(w(X'), c(X')) < h < \mathfrak{s}_X(w(X), c(X))$$

then  $cc' \preceq_w cc''$ .

That is, there are negative, slightly positive, and positive levels of epistemic utility such that it is always at least as good to add credences with the slightly positive epistemic utility as it is to add (perhaps far fewer) credences with the negative epistemic utility. Indeed, if we try to avoid the Paradox of Epistemic Utility for Credences by denying Benign Addition, we will see that we will also have to deny No Sadism, providing we accept the following innocuous principle:

**Addition** Suppose  $c, c', c''$  are credence functions defined on  $\mathcal{F}, \mathcal{F}', \mathcal{F}''$  respectively, where  $\mathcal{F}$  and  $\mathcal{F}'$  are disjoint and  $\mathcal{F}$  and  $\mathcal{F}''$  are disjoint. Then, if

- (i)  $|\mathcal{F}'| < |\mathcal{F}''|$ ;
- (ii) Every credence assigned by  $c, c'$ , or  $c''$  has positive epistemic utility;
- (iii) Each credence that  $c$  assigns has higher epistemic utility than each credence that  $c'$  assigns; and
- (iv) Each credence that  $c'$  assigns has higher epistemic utility than each credence that  $c''$  assigns;

then, if  $cc' \preceq_w c$ , then  $cc'' \preceq_w cc'$ .

The thought behind this principle is this. We start with your credence function  $c$ , all of whose credences have positive epistemic utility. First, we add some credences to  $c$  all of which have lower (but still positive) epistemic utility than all of the credences that  $c$  assigns. Now suppose that doing this doesn't improve your epistemic utility. Then, according to Addition, adding *more* credences all of which have *even lower* epistemic utility should not improve things either; indeed, it should make things at most as good as adding the smaller collection of better credences. Interestingly, Total and Average Epistemic Utilitarianism both satisfy Addition. Total Epistemic Utilitarianism satisfies it trivially because the antecedent is never satisfied: adding credences with positive epistemic utility will always improve  $c$ . Average Epistemic Utilitarianism satisfies it because, if the average epistemic utility of  $cc'$  is at most that of  $c$ , as the antecedent requires, then the average epistemic utility of  $cc''$  is at most that of  $cc'$ .

We can adapt the argument of Arrhenius (2000, Section 5-6) to show that if our epistemic utility ordering satisfies Partial Additivity, Transitivity, No Repugnance, and Addition, then it must violate No Sadism.

**Theorem 5** *There is no epistemic utility ordering  $\preceq_w$  that satisfies Partial Additivity, Transitivity, No Repugnance, Addition and No Sadism.*

Now, there's nothing inconsistent about denying both Benign Addition *and* No Sadism. For one thing, there are principles, such as Average Epistemic Utilitarianism and epistemic versions of the Variable Value Principles, that entail the negation of both of them. But, on the other hand, they both follow from the same more general and extremely plausible principle, namely, a slightly weakened epistemic analogue of what Michael Huemer calls the Modal Pareto Principle. Here's the ethical version (Huemer, 2008, 903):

**Modal Pareto Principle (ethical)** For any possible worlds  $x$  and  $y$ , if, from the standpoint of self-interest,  $x$  would rationally be preferred to  $y$  by every being who would exist *either* in  $x$  or in  $y$  or in both  $x$  and  $y$ , then  $x$  is better than  $y$  with respect to utility.

The (slightly weakened) epistemic version says the following:

**Modal Pareto Principle (epistemic)** For any credence functions  $c, c'$  defined on  $\mathcal{F}, \mathcal{F}'$  respectively, if  $c$  would be strictly better than  $c'$  from the standpoint of the local epistemic utility at  $w$  of every proposition that is *either* in  $\mathcal{F}$  or in  $\mathcal{F}'$  or in both  $\mathcal{F}$  and  $\mathcal{F}'$ , then  $c' \preceq_w c$ .

If we accept that, from the point of view of local epistemic utility at  $w$ , a proposition does better if it is assigned a credence with positive local epistemic utility than if it does not get assigned a credence at all, and it does better if it does not get assigned a credence at all than if it is assigned a credence with negative local epistemic utility, then the principle becomes:

**Modal Pareto Principle (epistemic)** Suppose  $c, c'$  are credence functions defined on  $\mathcal{F}, \mathcal{F}'$  respectively. And suppose

- (i) For each  $X$  in  $\mathcal{F} \cup \mathcal{F}'$ ,  $s_X(w(X), c'(X)) < s_X(w(X), c(X))$ ;
- (ii) For each  $X$  in  $\mathcal{F} - \mathcal{F}'$ ,  $0 < s_X(w(X), c(X))$ ;
- (ii) For each  $X$  in  $\mathcal{F}' - \mathcal{F}$ ,  $s_X(w(X), c'(X)) < 0$ .

Then  $c' \preceq_w c$ .

Benign Addition and No Sadism follow from this principle.

As Huemer says, it's hard to deny the ethical version of the Modal Pareto Principle. I submit that it's even more difficult to deny the epistemic version, since concerns about equality do not arise in that case. This, then, is our argument for Benign Addition: first, if we deny Benign Addition, we don't in fact escape the Paradox unless we also deny No Sadism or Addition. Furthermore, Benign Addition (like No Sadism) follows from the extremely plausible epistemic version of the Modal Pareto Principle.

### 3.2 Denying Partial Additivity

The second possible response to the Paradox of Epistemic Utility for Credences is to take it to be a new kind of argument against the sorts of trade-offs that are permitted by any account of epistemic utility for credences that endorses Partial Additivity. We already have some arguments against such trade-offs. The most sophisticated is due to Hilary Greaves (2013); it is based on her example, Epistemic Imps. Here's a version of this example:

**Epistemic Imps** Laia has very strong evidence in favour of the proposition *My car is in the garage*. Indeed, her evidence is strong enough to support certainty in that proposition — perhaps she's standing in front of the car looking at it. She also has credences in each of four propositions — *There's a set of car keys on the kitchen table*, *There's a set of car keys beside the television*, *There's a set of car keys in the bathroom*, *There's a set of car keys on the cat food box*. And she knows that her mischievous friend Àgata has all four sets of car keys and that she will determine

whether to put them in each of the places just mentioned depending on Laia's credence in *My car is in the garage*. If Laia's credence in that proposition is greater than 0, Ágata will toss a fair coin to decide whether to put a set of keys on the kitchen table, she'll toss another to decide whether to put a set of keys beside the television, and so on. So each of the four propositions about the sets of keys will have chance 0.5. On the other hand, if Laia has credence 0 in *My car is in the garage*, then Ágata will put a set of keys in each of the locations just mentioned for sure. So each of the four propositions about the sets of keys will have chance 1.

Thus, it seems that Laia has two options. She can have credence 1 in *My car is in the garage* and 0.5 in each of the propositions about the car keys. If she does that, she will respect her evidence concerning each proposition — her credence in *My car is in the garage* is justified by the evidence of her senses; and her credences in the other propositions match the known chances of those propositions (given her credence 1 in *My car is in the garage*). Or, she can have credence 0 in *My car is in the garage* and credence 1 in each of the propositions about the car keys. If she does that, she will fail to respect her evidence about the first proposition, but she will respect it about each of the others. However, at least if we use one of the quadratic scoring rules  $q_{\alpha,\lambda}$  introduced above, then the total epistemic utility of the former assignment of credences — namely, 1 in *My car is in the garage* and 0.5 in each of the others — is less at worlds at which Laia has those credences than the total epistemic utility of the latter assignment — namely, 0 in *My car is in the garage* and 1 in each of the others — is at worlds at which Laia has those credences. The point is that, although her credence in *My car is in the garage* is much less accurate in the second case, where she assigns it 0 instead of 1, her credences in the other propositions are much more accurate — they are assigned 1 and are true, rather than being assigned 0.5 and being either true or false.

Now, it might seem to follow from the preceding that accounts of epistemic utility that adhere to Partial Additivity must conclude that the former credences in this example are irrational, while the latter are rational. And indeed, that is precisely what Greaves claims. However, like Carr's objections to Total and Average Epistemic Utilitarianism above, this depends on a particular account of how we move from comparisons of the epistemic utility of credence functions at possible worlds to an account of irrationality. In particular, in Greaves' case, as she is well aware, it depends on an application of causal decision theory in epistemology in order to establish her conclusion. And such an application is controversial: it is accepted by Pettigrew (ta) and, with a tweak, by Joyce (ta); but it is denied by Konek & Levinstein (ms).

According to the response to the Paradox of Epistemic Utility for Credences that we are considering here, the Paradox amounts to an objection to the trade-offs endorsed by Partial Additivity; what's more, it's an objection that, unlike Greaves' argument from Epistemic Imps, does not rely on any particular account of how we derive facts about epistemic rationality from facts about epistemic utility — in particular, it does not rely on an application of causal decision theory in epistemology. The Paradox thus turns on the fact that Partial Additivity entails that there are pairs of credence functions,  $c^{++}$  and  $c^*$ , both defined on  $\mathcal{F}$ , where  $c^{++}$  is extremely accurate on  $q$  propositions and slightly accurate on the rest, while  $c^*$  is extremely inaccurate on  $n > q$  propositions, and slightly accurate on the rest, but such that  $c^{++} \prec_w c^*$ . Of course, the slightly accurate credences of  $c^{++}$  must be less accurate than the slightly accurate credences of  $c^*$ , but provided we are allowed to set the size of  $\mathcal{F}$  ourselves,

we can choose those slightly accurate credence of  $c^{++}$  and  $c^*$  so that they are as close together as we wish.

The problem with this response is that there are clear cases in which trade-offs exist and we wish to endorse them. Suppose I assign credence 0.6 to each of the arithmetical propositions  $1 + 1 = 2$  and  $2 + 2 = 4$ ; you assign 0.59 to  $1 + 1 = 2$  and 0.99 to  $2 + 2 = 4$ . It seems clear that you are doing better than I am. But of course that's because of a trade-off. Relative to me, you gain much greater accuracy in the latter arithmetical propositions, but you lose a little accuracy in the former. We naturally judge that the gain outweighs the loss, which is to say that we agree with the verdict of Partial Additivity and permit a trade-off.

Thus, if we are to make this response to the Paradox work, we must find a way to distinguish legitimate trade-offs, such as this one, from illegitimate ones — that is, we must say when the verdicts of Partial Additivity are correct and when they are not. Here is a proposal: No increase in total epistemic utility can ever compensate for moving from a credence function that assigns some credences with very high epistemic utility and none with very low epistemic utility to a credence function that assigns some credences with very low epistemic utility and none with very high epistemic utility. Thus, for instance,  $c^{++} \prec_w c^*$  doesn't hold, since  $c^{++}$  assigns some credences with very high epistemic utility and none with very low epistemic utility, while  $c^*$  assigns some credences with very low epistemic utility and none with very high. Thus, even though the total epistemic utility of  $c^*$  is greater than that of  $c^{++}$ , the latter is not worse than the former because this is a case that Partial Additivity gets wrong.

The ethical counterpart to this restriction on trade-offs could be motivated either by what Temkin (1987) calls Perfectionism or by what we might call Anti-Sadism. The former says that the existence of people living very good lives makes a population very good indeed; the latter says, amongst other things, that the existence of people living utterly miserable lives makes a population very bad indeed.

However, there are clear counterexamples to this proposal in the epistemic case. Consider the following example:

**More Birders** Maxine and Graham are out birding. There is a scarlet tanager in front of them. Maxine has high credence that there's a scarlet tanager in front of them while Graham has a very low credence in this proposition, having mistaken the bird for a northern cardinal. However, when we consider the credences that Maxine and Graham assign to the remaining 1 million propositions about which they have an opinion, both are only slightly inaccurate, but Graham is slightly more accurate than Maxine on each.

It seems natural to say that, in this case, Graham may well be doing better than Maxine. So the principle that we just stated cannot be true in full generality. There are at least some cases in which an increase in total epistemic utility compensates for the loss of all very highly accurate credences and the gain of some very inaccurate credences.

Perhaps, then, it is a question of quantity. After all, in the move from Maxine's credences to Graham's, we only lose one highly accurate credence and we only gain one highly inaccurate credence. Perhaps the correct principle is that we don't want to sacrifice *many* very accurate credences and gain *many* very inaccurate credences in order to gain greater total accuracy. However, we can use the example of Maxine and Graham to create counterexamples to this claim as well, providing we accept the following principle:

**Monotonic Combination** Suppose  $c_1, c_2$  are defined on  $\mathcal{F}$ , while  $c'_1, c'_2$  are defined on  $\mathcal{F}'$ . And suppose  $\mathcal{F}$  and  $\mathcal{F}'$  are disjoint. Then, if  $c_1 \preceq_w c_2$  and  $c'_1 \prec_w c'_2$ , then  $c_1 c'_1 \prec_w c_2 c'_2$ .

Note that there was nothing special about the propositions on which Maxine's and Graham's credences were defined that made our original counterexample plausible. Thus, we might easily assume that there are many disjoint sets of proposition  $\mathcal{F}_1, \dots, \mathcal{F}_n$  together with credence functions  $b_1, \dots, b_n$  and  $c_1, \dots, c_n$  defined on  $\mathcal{F}_1, \dots, \mathcal{F}_n$  respectively such that each  $b_i$  is like Maxine's credence function — that is, it assigns one very accurate credence and the rest of the credences it assigns are slightly accurate — and each  $c_i$  is like Graham's — that is, it assigns one very inaccurate credence and the rest of the credences it assigns are slightly accurate, and slightly more accurate than Maxine's corresponding credences. Then, by analogy with Maxine and Graham, we might assume that  $b_i \prec_w c_i$ . And so, by Monotonic Combination, we obtain  $b_1 \dots b_n \prec_w c_1 \dots c_n$ . And this gives a counterexample to the proposal we're considering, which says that no gain in total epistemic utility can ever compensate for the loss of many very accurate credences and the gain of many very inaccurate credences.

So it seems that the sorts of trade-off that Partial Additivity endorses and that lead to  $c^{++} \prec_w c^*$  are exactly the sorts of trade-offs that we either countenance intuitively or that follow from trade-offs that we countenance intuitively, together with plausible further principles, such as Monotonic Combination.

### 3.3 Denying Transitivity

In this section, we consider whether we might escape the Paradox of Epistemic Utility for Credences by denying that the relation  $\preceq_w$  is transitive. In the ethical case, there are standardly two ways to argue for this. Let's begin with the first, where we argue against the transitivity of the better-than relation between populations by creating an apparent counterexample. This consists of a very long sequence of populations  $A_1, \dots, A_{1,000,000}$ . The first population,  $A_1$ , contains 2 people in terrible agony. The second,  $A_2$ , contains 4 people in very slightly less terribly agony than those in  $A_1$ .  $A_3$  contains 8 people in very slightly less terribly agony than those in  $A_3$ . And so on. The final population in the sequence,  $A_{1,000,000}$ , contains  $2^{1,000,000}$  people with a very mild headache (Rachels, 1998). It is claimed that  $A_1$  is better than  $A_2$ ,  $A_2$  is better than  $A_3$ ,  $A_3$  is better than  $A_4$ , ...  $A_{999,999}$  is better than  $A_{1,000,000}$ . After all, doubling the number of people in a population while reducing the level of suffering they each endure only by a tiny amount makes a situation worse. However, it is also claimed that  $A_{1,000,000}$  is better than  $A_1$ . After all, a population all of whose members are in terribly agony is always worse than a population all of whose members merely have a mild headache. Thus, if transitivity holds,  $A_1$  is better than itself, which is a contradiction: if  $A_1$  is better than  $A_1$ , then by definition  $A_1$  is at least as good as  $A_1$  and  $A_1$  is not at least as good as  $A_1$ .

Let's now consider the epistemic analogue. This consists of a sequence of credence functions  $c_1, \dots, c_{1,000,000}$ . The first assigns only credences with extremely low epistemic utility at  $w$ ; each credence function thereafter is defined on twice as many propositions as the previous one; and each assigns credences that have only very slightly more epistemic utility at  $w$  than those assigned by the previous one; and the final one assigns credences that all have only very slightly negative epistemic utility at  $w$ . As in the ethical case, we would have

$c_1 \succ_w c_2 \succ_w \dots \succ_w c_{1,000,000} \succ_w c_1$ . And again this means that Transitivity would entail  $c_1 \succ_w c_1$ , which is logically impossible. The problem with the putative counterexample in this case is that the intuition that the final credence function,  $c_{1,000,000}$ , in the sequence is epistemically better than the first,  $c_1$ , is not so strong. In the ethical case, we intuitively abhor any situation in which people experience excruciating agony — we are prepared to tolerate a great deal in order to avoid such a situation. In the epistemic case, on the other hand, the intuitive pull is not so strong — as we saw in More Birders above. We do not have anything like the same abhorrence of the credence function  $c_1$  as we have of the population  $A_1$ . The reason, it seems to me, is that the intuitive pull of Anti-Sadism is dramatically stronger in the ethical case than in the epistemic case: we abhor populations that contain great suffering; we do not so strongly disvalue credence functions that assign credences with very low epistemic utility.

This makes the epistemic case particularly vulnerable to objections that target the very possibility of epistemic utility cycles, however formed, such as the cycle  $c_1 \succ_w c_2 \succ_w \dots \succ_w c_{1,000,000} \succ_w c_1$  from above. We will see such an objection below. I take this objection to be decisive. Thus, I take the putative cycle given above not to be genuine. Indeed, I take it that  $c_1$  genuinely is better than  $c_{1,000,000}$ , just as transitivity would suggest: it is better to be badly wrong on just a few propositions than slightly wrong on an enormous number of them.

The second way to argue against Transitivity in the ethical case is endorsed by Larry Temkin (1987). Temkin argues that the better-than relation between populations might be *comparative*. The idea is this: There are a number of different dimensions on which we might compare two items — two populations, perhaps, or two credence functions — that will jointly determine an overall comparison of their goodness. These comparisons along different dimensions must then be summarized into an overall all-things-considered better-than relation. If we always weight the comparisons along the different dimensions in the same way, no matter which two items we are comparing, then the all-things-considered better-than relation will be transitive. However, Temkin claims that, at least in the case of the better-than relation between populations, those weightings change depending on the items we're comparing. For instance, when we compare two populations that contain the same number of people, we may give considerable weight to the utilities of the worst-off. On the other hand, when we compare two populations that differ only by the addition of people who are living lives worth living, we do not give any weight to the utilities of the worst-off. In this way, the better-than relation is comparative, and we no longer have any reason to think that it is transitive.

Let's see how this might work in the epistemic case. As with the ethical case, there are many dimensions along which we compare credence functions. We compare them by looking at whether they contain any credences with very high epistemic utility — we might call this the *perfectionist* dimension, following Temkin. We compare them by looking at whether they contain any credences with very low epistemic utility — we might call this the *anti-sadist* dimension, by analogy with the ethical case. And we compare them by looking at their total epistemic utility. Now let's consider the credence functions that feature in the Paradox of Epistemic Utility for Credences.  $c^{++}$  is at least as good as  $c^+$  on all dimensions of comparison.  $c^*$  is better than  $c^{++}$  on the dimension of total epistemic utility and worse on the perfectionist and anti-sadist dimensions.  $c^*$  is better than  $c^+$  on the total epistemic utility dimension, but worse on the perfectionist and anti-sadist dimension. Now, if we weight the three dimensions in the same way when we compare  $c^*$  and  $c^+$  and when we compare  $c^*$  and  $c^{++}$ , then we must have  $c^+ \prec_w c^*$  iff  $c^{++} \prec_w c^*$ . But if we have that, then we have  $c^+ \prec_w c^*$ ,

since Partial Additivity gives  $c^{++} \prec_w c^*$ ; and  $c^+ \prec_w c^*$  is precisely the consequent of the instance of Transitivity that the Paradox requires. So a contradiction follows. Thus, we must weight these three dimensions differently depending on which comparisons we're making, just as Temkin suggests.

By Partial Additivity, it follows that we give no weight to the perfectionist and anti-sadist dimensions if we are comparing credence functions on the same set of propositions — just as, in the ethical case, we pay no attention to the plight of the worst-off when we are comparing a population and an extension of it obtained by merely adding lives worth living. However, when we come to compare  $c^+$  and  $c^*$ , we do not weight the total epistemic utility dimension above all others. We also take account of the perfectionist and anti-sadist dimensions. And, doing that, we conclude that  $c^+ \succeq_w c^*$ .

Of course, once we have  $c^* \preceq_w c^+$ , we have an epistemic utility cycle  $c^+ \preceq_w c^{++} \prec_w c^* \preceq_w c^+$ . Unlike the cycle  $c_1 \succ_w c_w \succ_w \dots \succ_w c_{1,000,000} \succ_w c_1$ , this is not a strict cycle — that is, its inequalities are not all strict. But, as we will see, both fall to the same objection. Before we consider that objection, however, we consider another objection to intransitive better-than orderings: this is intended to apply whether or not these orderings issue in cycles.

The objection is due to John Broome (2004, Section 4.1), who claims that it is an analytic truth that the better-than relation is transitive. His argument targets particularly the better-than relation between possible lives and the better-than relation between possible populations. However, if it is successful in these cases, it seems that it should be successful quite generally; in particular, it should establish that the epistemic better-than relation between credence functions is also transitive — that is, it should establish that our Transitivity axiom is an analytic truth.

Broome claims that the better-than relation stands to the adjective *good* as the taller-than relation stands to the adjective *tall* and the younger-than relation stands to the adjective *young*. To say that population *A* is better than population *B* is to say that the former is *more good than* the latter. And, according to Broome, the correct way to understand these sorts of relation is as positing degrees of the property that the adjective in question picks out — degrees of height, youth, goodness — and as saying that both relata have a degree of the given property and that the degree of the first relatum is greater than the degree of the second relatum. This understanding of the more-*F*-than (or *Fer*-than) relations for a predicate *F* is also accepted by linguists, who incorporate the preceding analysis, complete with its ontology of degrees of *F*-ness, into their semantics for such relations (Kennedy, 2001). The transitivity of the *Fer*-than relation is then taken to follow from the transitivity of the greater-than relation on the set of possible degrees that are ascribed to the relata of the *Fer*-than relation by the semantics. Thus, the transitivity of *Fer*-than — and, in particular, the various better-than relations between possible populations, lives, and credence functions — is an analytic truth.

The problem with this argument is that it over-generates. It relies on the assumption that the greater-than relation on the degrees of *F*-ness posited by the semantics of the *Fer*-than relation is itself transitive. Whence comes our confidence in this? Surely from the analogy with the greater-than relation on the set of real numbers, which we often use to provide those degrees — degrees of tallness, heaviness, largeness, hardness, richness, etc. are all given by real numbers relative to a particular unit. However, if that's right, we should be equally confident that the greater-than relation on the set of degrees is *complete* — that is, for any two degrees, the first is greater than the second, the second is greater than the first, or they are equal. Thus, we can conclude that it is an analytic truth not only that the *Fer*-than relation



is transitive for any  $F$ , but also that it is complete: that is, for any two possible items  $a$  and  $b$  such that it is not a category mistake to ascribe  $F$ -ness to these items, it is the case that  $a$  is  $F$ er than  $b$ ,  $b$  is  $F$ er than  $a$ , or  $a$  is exactly as  $F$  as  $b$ . Thus, it is analytically false to say, as Ruth Chang (2002, 2005) does, that Mozart is not more creative than Michelangelo, Michelangelo is not more creative than Mozart, and Mozart and Michelangelo are not equally creative. But that seems wrong. Chang's claim is extremely plausible, and indeed certainly more plausible than the highly theoretical claim that the greater-than relation on the degrees of  $F$ -ness is always transitive, or indeed the more basic but still highly theoretical claim that the semantics of the  $F$ er-than relation always involves ascribing degrees of goodness. So Broome's argument fails to convince.

Nonetheless, as I intimated earlier, the Temkin-style response fails all the same. The problem is one that haunts any account that posits cyclical preferences. Let us say that an *epistemic utility cycle* is a sequence of credence function  $c_1, \dots, c_n$  with the following properties:

- (i)  $c_i \succeq_w c_{i+1}$  for all  $i = 1, 2, \dots, n-1$ , and  $c_n \succeq_w c_1$ ;
- (ii)  $c_i \succ_w c_{i+1}$  for some  $i = 1, 2, \dots, n-1$ , or  $c_n \succ_w c_1$ .

Thus,  $c_1 \succ_w c_2 \succ_w \dots \succ_w c_{1,000,000} \succ_w c_1$  and  $c^* \preceq_w c^+ \preceq_w c^{++} \prec_w c^*$  are both epistemic utility cycles.

Now, so far, we have been considering an epistemic utility ordering on credence functions: these have been the items whose epistemic goodness we have compared using the ordering. However, it is natural to think that we can extend the ordering to *mixtures* of these items. Given a set of credence functions  $c_1, \dots, c_n$ , a mixture of those credence functions is an option that results in the agent adopting  $c_1$  with some probability,  $c_2$  with some probability,  $\dots$ ,  $c_n$  with some probability. This way of creating new options out of old is familiar from decision theory and game theory, where the old options are the acts or the strategies, respectively, and the new options are the mixed acts and mixed strategies. Here's some notation so that we can refer to these mixed options composed out of credence functions. Given credence functions  $c_1, \dots, c_n$  and a probability distribution  $p_1, \dots, p_n$  (so  $\sum_{i=1}^n p_i = 1$ ), we write  $\{p_1, c_1; \dots; p_n, c_n\}$  for the mixed option that gives credence function  $c_i$  with probability  $p_i$ . Thus,  $\{0.75, c; 0.25, c'\}$  is the mixed option that gives  $c$  with probability 0.75 and  $c'$  with probability 0.25. Now suppose we include all possible mixtures of credence functions in the domain of our epistemic utility ordering  $\preceq_w$ . What principles might govern this expanded ordering? Here's a principle that seems uncontroversial: If  $c_1 \succ_w c_2$  and  $c'_1 \succ_w c'_2$ , then  $\{p, c_1; 1-p, c'_1\} \succ_w \{p, c_2; 1-p, c'_2\}$  for any  $0 \leq p \leq 1$ . And here's another: If  $c \succ_w c'$ , then  $\{1, c\} \succ_w \{p, c; 1-p, c'\}$ , for all  $0 \leq p < 1$ . Now, each of these putative principles is a particular instance of a more general principle, which is also extremely plausible:

**Stochastic Dominance** If  $c_1, \dots, c_n, c'_1, \dots, c'_n$  are credence functions and

- (i)  $c_i \succeq_w c'_i$  for all  $i = 1, \dots, n$ ;
- (ii)  $c_i \succ_w c'_i$  for some  $i = 1, \dots, n$ ;

then

$$\{p_1, c_1; \dots; p_n, c_n\} \succ_w \{p_1, c'_1; \dots; p_n, c'_n\}$$

for any  $0 < p_1, \dots, p_n < 1$  with  $\sum_{i=1}^n p_i = 1$ .

However, the problem is that Stochastic Dominance rules out cycles.<sup>11</sup>

**Theorem 6** *There is no epistemic utility ordering  $\preceq_w$  on mixed options that contains an epistemic utility cycle amongst the unmixed options and that satisfies Stochastic Dominance.*

Epistemic utility cycles, then, are ruled out by general principles that are more plausible than the considerations that lead us to posit the cycles in the first place. So, it seems, we should not try to escape the Paradox by denying Transitivity.

### 3.4 Denying No Repugnance

In previous sections, we have considered trying to escape the Paradox of Epistemic Utility for Credences by rejecting Partial Additivity or Transitivity or Benign Addition. In this final section, we consider rejecting No Repugnance. That is, we consider accepting the epistemic version of the Very Repugnant Conclusion. For the veritist, this means that we consider accepting that, for any credence function that assigns only very accurate credences, and any credence function that assigns only very inaccurate credences, we can extend the latter by adding only very slightly accurate credences so that it ends up better than the former.

Now, recall that the epistemic version of the Very Repugnant Conclusion is a consequence of Total Epistemic Utilitarianism. But the implication doesn't run in the other direction — rejecting No Repugnance and accepting the Very Repugnant Conclusion does not commit us to setting the epistemic utility ordering of two credence functions in line with their total epistemic utilities. Nonetheless, if we accept two of the principles that give rise to the Very Repugnant Conclusion via the Paradox of Epistemic Utility for Credences — namely, Benign Addition and Transitivity — together with a slight strengthening of the third — Partial Additivity — and two new principles — Malign Addition and Order Continuity, stated below — then we can infer Total Epistemic Utilitarianism.

**Malign Addition** Suppose  $c$  and  $c'$  are credence functions on  $\mathcal{F}$  and  $\mathcal{F}'$  respectively. And suppose  $\mathcal{F} \subseteq \mathcal{F}'$ . Then, if

- (i) For each proposition  $X$  that is in  $\mathcal{F}'$  and also in  $\mathcal{F}$ , the credence that  $c'$  assigns to  $X$  has lower epistemic utility at  $w$  than the credence that  $c$  assigns to  $X$ , and
- (ii) For each proposition  $X$  that is in  $\mathcal{F}'$  but not also in  $\mathcal{F}$ , the credence that  $c'$  assigns to  $X$  has negative epistemic utility at  $w$ ,

then  $c \succeq_w c'$ .

Like Benign Addition, Malign Addition follows from the epistemic version of the Modal Pareto Principle stated in Section 3.1 above.

**Order Continuity** Suppose that  $c_1, c_2, \dots$  is an infinite sequence of credence functions on  $\mathcal{F}$  and  $c$  is also a credence function on  $\mathcal{F}$ . We say that  $c_n$  approaches  $c$  as  $n$  tends to infinity (and write  $\lim_{n \rightarrow \infty} c_n = c$ ) if

$$(\forall \varepsilon > 0)(\exists N)(\forall M > N)(\forall X \in \mathcal{F}) [|\mathfrak{s}_X(w(X), c_M(X)) - \mathfrak{s}_X(w(X), c(X))| < \varepsilon]$$

Then, suppose  $c^*$  is also a credence function on  $\mathcal{F}$ .

---

<sup>11</sup>The result is inspired by (Gustafsson, 2010), which is concerned with producing money-pump arguments against non-strict preference cycles.

- If  $\lim_{n \rightarrow \infty} c_n = c$  and  $c_n \preceq_w c^*$  for all  $n$ , then  $c \preceq_w c^*$ .
- If  $\lim_{n \rightarrow \infty} c_n = c$  and  $c_n \succeq_w c^*$  for all  $n$ , then  $c \succeq_w c^*$ .

This says that, if each member of a sequence of credence functions whose local epistemic utilities come arbitrarily close to those of  $c$  is at most as good as  $c^*$ , then  $c$  is at most as good as  $c^*$ ; and if each member of a sequence of credence functions whose local epistemic utilities come arbitrarily close to those of  $c$  is at least as good as  $c^*$ , then  $c$  is at least as good as  $c^*$ . That is, you can't have a sequence of credence functions all no better than  $c^*$  but approaching a limit that is better than  $c^*$ ; and you can't have a sequence of credence functions all no worse than  $c^*$  but approaching a limit that is worse than  $c^*$ .

Here is the slight strengthening of Partial Additivity. It drops the requirement that  $c$  and  $c'$  are defined on the same *set* of propositions, and replaces it with the weaker requirement that  $c$  and  $c'$  are defined on the same *number* of propositions.

**Partial Additivity<sup>+</sup>** If  $c, c'$  are credence functions defined on  $\mathcal{F}, \mathcal{F}'$  respectively, and  $|\mathcal{F}| = |\mathcal{F}'|$ , then

$$c \preceq_w c' \text{ iff } \sum_{X \in \mathcal{F}} s_X(w(X), c(X)) \leq \sum_{X' \in \mathcal{F}'} s_{X'}(w(X'), c'(X'))$$

We now have the following theorem:

**Theorem 7** *Benign Addition + Malign Addition + Partial Additivity<sup>+</sup> + Transitivity + Order Continuity + Utility Continuity + Truth-Directedness  $\Rightarrow$  Total Epistemic Utilitarianism*

Thus, if we deny No Repugnance because we accept the other principles that give rise to the Paradox of Epistemic Utility for Credences, we are very close to accepting Total Epistemic Utilitarianism.

Why should we accept the other principles and deny No Repugnance? That is, why should we accept the epistemic version of the Very Repugnant Conclusion — and perhaps further Total Epistemic Utilitarianism, from which the Very Repugnant Conclusion follows — rather than rejecting Transitivity, Benign Addition, or Partial Additivity? There are three approaches we might adopt when answering this question.

The first is a very crude cost-benefit analysis: we weigh the intuitive pull of each of the four principles that make up the Paradox of Epistemic Utility for Credences against the others; and we claim that No Repugnance has least intuitive support. I believe this analysis is correct, but the strength of the intuitive support that a principle enjoys is notoriously difficult to measure, and it is notoriously difficult even to make comparisons in terms of intuitive support. Furthermore, these comparisons may vary from one person to another.

Our second strategy is a more sophisticated cost-benefit analysis. We add to our intuitive reactions to these principles also the considerations we have adduced in favour of the three of them other than No Repugnance in the preceding sections. For instance, we might note that Benign Addition follows from the very plausible epistemic version of the Modal Pareto Principle; and we might observe that the standard objections to Transitivity issue in utility cycles, which are ruled out by the very plausible principle of Stochastic Dominance.

This argument in favour of accepting the other axioms and consequently denying No Repugnance is much stronger than the first. But I wish to strengthen it yet further by combining it with a third sort of argument, namely, a debunking argument. That is, as well as appealing to the positive intuitions and arguments in favour of the other principles, I wish to give a

negative argument that targets the intuition in favour of No Repugnance — I want to give a debunking explanation for that intuition that thereby undermines it.<sup>12</sup> Recall the example of Daphne and Phoebe from Section 2.3 above. In it, Phoebe has 1 million credences, each of which is extremely accurate; Daphne, in contrast, has 1 trillion credences, each of which is extremely inaccurate, as well as many other credences, each of which is only slightly accurate. The Very Repugnant Conclusion claims that, in such a situation, Daphne's credal state might be better than Phoebe's. Our intuition, it seems, ranks Phoebe's credal state above Daphne's. I claim that our intuitions get this wrong. Here is my debunking explanation for this error: I claim that we intuitively reject the correct ordering of Daphne above Phoebe — to the extent that we do — because we make our judgment about the *epistemic* utility of their respective credal states by appealing directly to our judgment concerning the *pragmatic* utilities of those states. That is, when asked to rank Phoebe and Daphne's credal states in terms of their epistemic goodness, we defer to our ranking of them in terms of their pragmatic goodness. Now, what is the pragmatic goodness or utility of a certain set of credences? Well, since credences, together with utilities, give rise to our rational decisions, the pragmatic utility of our credal state is the pragmatic utility of the outcomes of the decisions that we use it to make — this is the insight of so-called 'Dutch Book' arguments for Probabilism and Conditionalization. Now it seems correct to say that Phoebe's credences will guide her better in the world than Daphne's will. For a huge number of propositions, Daphne will go wrong in very many decisions that turn on her attitudes to those propositions. And for many others, she will not go right much more often than if she were simply to toss a coin. Phoebe, in contrast, will go right very often when the decision turns on a proposition to which she assigns a credence. For these reasons, I submit, we judge that Phoebe's credences are better than Daphne's in terms of their pragmatic utility. I also claim that, when we are asked which, if either, is better in terms of their epistemic utility, we defer to this judgment. Why do we do that? For two reasons, I think. First, doing so is a reliable method by which to make such a judgment. Since, in normal cases, we get what we want just in case we accurately represent the world, high pragmatic utility is a reliable indicator of high epistemic utility, low pragmatic utility is a reliable indicator of low epistemic utility, and so on. Second, it is cognitively easier to assess a credal state for its pragmatic utility than for its epistemic utility. This is because it is easier cognitively to simulate the concrete consequences of having a given set of credences and assess the pragmatic utility of those consequences than it is to calculate or estimate the more abstract quantity of epistemic utility.

This, then, debunks the intuition in favour of No Repugnance. By doing so, we clear the way for denying No Repugnance and accepting the epistemic version of the Very Repugnant Conclusion. However, as we saw in Theorem 7 above, if we accept the remaining three axioms that make up the Paradox, together with some further related plausible principles, it is not just the Very Repugnant Conclusion that we must accept, but also Total Epistemic Utilitarianism. And, as we saw in Section 1.2, Jennifer Carr has identified some unpalatable consequences of Total Epistemic Utilitarianism when it is combined with natural decision-theoretic principles. Recall: Theorem 2 teaches us that, if we accept Total Epistemic Utilitarianism, we face a dilemma. Either (a) for any non-zero credence, there is a proposition whose epistemically neutral credence is lower than that, or (b) there is a credence  $r$  and an open proposition  $X$  such that any credence function that assigns  $r$  to  $X$  is strictly dominated. As we saw in Section 1.2 above, the former flies in the face of veritism, while the latter seems to

---

<sup>12</sup>This approach was inspired by a comment from an anonymous referee for this journal.

pose problems for accuracy-first arguments for Probabilism, as well as related arguments for other credal norms, such as Conditionalization and the Principal Principle. I wish to argue that, while (a) is unsupportable, (b) is acceptable and poses no problem for the accuracy-first arguments in question.

To introduce the idea behind my argument, compare the following three cases:

- *Memory* I expect that, if I were to remember more past experiences than I do, and if I were to remember those I do more accurately than I do, I would have more accurate credences than I do.
- *Evidence* I expect that, if I were to collect more evidence than I currently have, I would have more accurate credences than I do.
- *Opinions* If option (b) above is true, then there is a credence  $r$  and a proposition  $X$  such that if I currently assign  $r$  to  $X$ , then there is some *a priori* knowable way to alter my opinion set and credence function such that, if I were to undertake this alteration, I would be guaranteed to have more accurate credences than I do.

In the first two cases, we tend not to judge me irrational on the basis that there's something epistemic I could do that I can expect to be better for me, epistemically speaking — as Williamson (1998, 98) says, "Forgetting is not irrational; it is just unfortunate". Perhaps, then, for the same reason, we should not judge me irrational in the third case, where there's something epistemic I could do that is guaranteed to improve my credal state from the purely epistemic point of view.

I think that's right. But why? Here's my proposed explanation. When we give certain norms that govern credences, such as Probabilism or Conditionalization, we are clearly not claiming that having non-probabilistic credences or updating by a rule other than conditionalization is *always all-things-considered irrational*. After all, I might be in a situation in which violating Probabilism is the only way to save the lives of my loved ones; and in such a situation, it is all-things-considered rational to save the lives of my loved ones at the expense of the coherence of my credences. Rather, we are claiming that norms such as Probabilism and Conditionalization are *epistemic* norms that govern *assignments of credences*. They are not *all-things-considered* norms that govern *an agent's actions in general*. Seeing them as epistemic norms that govern assignments of credences has two consequences. First, the fact that we are focussing on *purely epistemic* rationality entails that, when we apply the principles of decision theory to derive facts about the rationality of the options we are assessing from facts about their value, it is their purely epistemic value that we will consider — we will simply ignore their practical or moral or aesthetic value. Second, the fact that we are focussing on the rationality of *assignments of credences* entails that certain other sorts of epistemic actions — such as collecting more evidence or improving our memory to provide more support for our credences, or altering the opinion set to whose members we assign those credences — are not included as available options in the decision problem we use to assess the rationality of the assignments. The point is that, when we focus on the all-things-considered rationality of an option, we assess it using a decision problem with two features: first, the utility function measures all-things-considered (or overall) value, combining epistemic value with practical value and moral value and aesthetic value and so on; second, the set of available options includes, roughly speaking, anything that it is physically possible for the agent to do. When, instead, we narrow down our focus to what is practically rational or what is epistemically

rational, we still use decision theory to carry out our assessment, but the decision problem we use differs in both of these features: we restrict the sort of value measured by the utility function; and we restrict the options available to the agent. When it is the epistemic rationality of assignments of credences that we are using the decision problem to assess, we allow our utility function to measure only epistemic value; and we consider as available options only other assignments of credences to the same propositions.

One attractive feature of this account is that it explains why we sometimes do, nonetheless, impugn the rationality of an agent who doesn't take free evidence, or who refuses a free pill that will improve her memory with no side effects, or who doesn't make a cost-free change to her opinion set that is guaranteed to improve her epistemic utility. The explanation is this: in these cases, we have shifted from evaluating the epistemic rationality of an assignment of credences; instead, we are evaluating the overall or all-things-considered rationality of an agent's total doxastic state, which includes her evidence (including the evidence she obtains from her memory) and her opinion set. As a result, we expand the space of available options in the decision problem that we use to evaluate the rationality of the agent's current state. There are now new options, such as altering your opinion set or collecting new evidence. Now, it might seem that, by expanding the set of options in this way, we will return to the situation we had above, where each option is dominated. However much evidence I gather, I could have gathered more; and I expect that doing so would have increased by epistemic utility. This would be a problem if we were still evaluating these options using only their epistemic utility. But we are now considering also their practical utility, and of course gathering new evidence often comes at a cost — it takes time and energy to investigate the world — as does altering one's opinion set — it takes time and energy to formulate any new propositions you introduce as well as space to retain them, and it takes time and energy to ignore any old propositions you wish to jettison. Thus, while we sometimes criticise someone who refuses free evidence or who fails to make a cost-free change to her opinion set that is guaranteed to increase her epistemic utility, we recognise that there are other times when sticking with her current evidence and her current opinion set maximizes (expected) all-things-considered utility, which is the relevant utility when making this assessment of the agent's total doxastic state, rather than simply her assignment of credences.

Thus, in conclusion, while Total Epistemic Utilitarianism might entail that many probabilistic credence functions are strictly dominated with respect to accuracy, this does not undermine the accuracy-first arguments that have been given for Probabilism, Conditionalization, and so on. After all, these norms are epistemic norms that govern assignments of credences: this means that the only available options in the decision problem we use to evaluate a given assignment of credences includes other assignments *to the same set of propositions*. And, once we recognise that, Joyce's argument goes through. After all, provided our epistemic utility function has certain properties that we outlined in Section 1.1.3, and providing our epistemic utility ordering satisfies Total Epistemic Utilitarianism, each non-probabilistic credence function is strictly dominated by a probabilistic credence function defined on the same opinion set; and no probabilistic credence function is strictly dominated by any credence function defined on the same opinion set (Joyce, 1998, 2009; Predd et al., 2009; Pettigrew, 2016).

## 4 Conclusion

To summarize: I propose that our epistemic Theory X is Total Epistemic Utilitarianism. It follows from Benign Addition, Malign Addition, Transitivity, Partial Additivity, and Order and Utility Continuity. And, while it entails the Very Repugnant Conclusion (that is, the negation of No Repugnance), that is independently the conclusion of Benign Addition, Transitivity, and Partial Additivity, each of which it is more costly to deny than No Repugnance. The most troubling feature of Total Epistemic Utilitarianism for those of us who wish to preserve the accuracy-first arguments for Probabilism, Conditionalization, etc. is the consequence identified by Theorem 2. However, as we saw in the previous section, it is possible to preserve these accuracy arguments even in the presence of the most worrying possible consequences of that theorem. The population ethics for credence, therefore, is simply the epistemic analogue of total utilitarianism in ethics.

As I mentioned at the beginning, this paper is intended as a case study in the population ethics of doxastic states. We have focussed only on credences; and we have taken their epistemic utility to be exhausted by their accuracy. Thus, future work might generalise this framework along these two dimensions. We might consider alternative sorts of doxastic state: full beliefs, primitive conditional credences, imprecise credences, comparative credences, etc. And we might consider a less austere roster of sources of epistemic value: matching evidential probabilities, constituting understanding, constituting knowledge, formed by a reliable process, etc. There is no reason to think that the same problems arise nor that the same solution will be best in all of these different cases.

## Appendix: Proofs of theorems

**Proposition 1** Suppose that  $\mathfrak{s}$  is truth-directed and continuous. Then every proposition has an epistemically neutral credence.

*Proof.* Let  $X$  be a proposition. Since  $\mathfrak{s}$  is truth-directed and continuous  $\mathfrak{s}_X(1, x)$  is a continuous, strictly increasing function of  $x$  and  $\mathfrak{s}_X(1, x)$  is a continuous, strictly decreasing function of  $x$ . Moreover,  $\mathfrak{s}(1, 0) < \mathfrak{s}(0, 0)$  but  $\mathfrak{s}(1, 1) > \mathfrak{s}(0, 1)$ . So the functions  $\mathfrak{s}(1, x)$  and  $\mathfrak{s}(0, x)$  must intersect at exactly one point between 0 and 1 — let us call the point of their intersection  $r_X$ . So  $\mathfrak{s}(1, r_X) = \mathfrak{s}(0, r_X)$ , as required.  $\square$

**Theorem 2** Suppose that  $\mathfrak{s}$  is truth-directed, strictly proper, and continuous. And suppose that the epistemic utility ordering  $\preceq_w$  is governed by Total Epistemic Utilitarianism. Then either:

- (a) For all  $\varepsilon > 0$ , there is a proposition  $X$  such that  $0 \leq r_X < \varepsilon$ ;
- OR
- (b) There is a credence  $x$  and an open proposition  $X$  such that every credence function that assigns  $x$  to  $X$  is strictly dominated.

*Proof.* First, note that one of the following two must hold:

- (i) There is an open proposition  $X$  such that the epistemically neutral credence  $r_X$  for  $X$  is guaranteed to have negative epistemic utility — that is,  $\mathfrak{s}_X(w(X), r_X) = \mathfrak{s}_X(w'(X), r_X) < 0$ , for all  $w, w'$ .

- (ii) For all open propositions  $X$ , the epistemically neutral credence  $r_X$  for  $X$  is guaranteed to have non-negative epistemic utility — that is,  $\mathfrak{s}_X(w(X), r_X) = \mathfrak{s}_X(w'(X), r_X) \geq 0$  for all  $w, w'$ .

If (i), then (b) follows from Total Epistemic Utilitarianism. After all, suppose  $c$  is defined on an opinion set  $\mathcal{F}$  that contains  $X$ . And suppose  $c(X) = r_X$ . Then define  $c^-$  on the opinion set  $\mathcal{F}^- = \mathcal{F} - \{X\}$  so that  $c^-(Y) = c(Y)$  for all  $Y$  in  $\mathcal{F}^-$ . Then

$$\begin{aligned} \sum_{Y \in \mathcal{F}} \mathfrak{s}_Y(w(Y), c(Y)) &= \mathfrak{s}_X(w(X), c(X)) + \sum_{Y \in \mathcal{F}^-} \mathfrak{s}_Y(w(Y), c(Y)) \\ &= \mathfrak{s}_X(w(X), r_X) + \sum_{Y \in \mathcal{F}^-} \mathfrak{s}_Y(w(Y), c(Y)) \\ &< \sum_{Y \in \mathcal{F}^-} \mathfrak{s}_Y(w(Y), c(Y)) \\ &= \sum_{Y \in \mathcal{F}^-} \mathfrak{s}(w(Y), c^-(Y)) \end{aligned}$$

So, by Total Epistemic Utilitarianism,  $c \prec_w c^*$  for all worlds  $w$ .

Thus, we assume (ii). What's more, we assume that (a) is false; we will then show that (b) follows. In fact, we show that, if (ii) is true and (a) is false, then *every* credence function is strictly dominated; thus, there is certainly a credence  $x$  and an open proposition  $X$  such that every credence function that assigns  $x$  to  $X$  is strictly dominated.

If (ii) is true and (a) is false, and  $c$  is a credence function on  $\mathcal{F}$ , then there are mutually exclusive open propositions  $X_1, \dots, X_n$  that are not included in  $\mathcal{F}$  such that the sum of their epistemically neutral credences exceeds 1 — that is,  $\sum_{i=1}^n r_{X_i} > 1$ . Then we will define  $c'$  on the opinion set  $\mathcal{F}' = \mathcal{F} \cup \{X_1, \dots, X_n\}$  that is obtained by extending  $\mathcal{F}$  to include  $X_1, \dots, X_n$ . We will let  $c'$  agree with  $c$  on  $\mathcal{F}$ , so that  $c'(X) = c(X)$  for all  $X$  in  $\mathcal{F}$ . And we will let  $c'$  assign to each  $X_i$  its epistemically neutral credence  $r_{X_i}$ , so that  $c'(X_i) = r_{X_i}$  for all  $i = 1, \dots, n$ . Then, for all worlds  $w$ ,

$$\begin{aligned} \sum_{X' \in \mathcal{F}'} \mathfrak{s}_{X'}(w(X'), c'(X')) &= \sum_{X \in \mathcal{F}} \mathfrak{s}_X(w(X), c(X)) + \sum_{i=1}^n \mathfrak{s}_{X_i}(w(X_i), r_{X_i}) \\ &\geq \sum_{X \in \mathcal{F}} \mathfrak{s}_X(w(X), c(X)) \end{aligned}$$

since  $\mathfrak{s}(w(X_i), r_{X_i}) = \mathfrak{s}(w'(X_i), r_{X_i}) \geq 0$  for all worlds  $w, w'$ . Now, notice that  $c'$  is non-probabilistic. After all, it assigns credences to a set of mutually exclusive propositions whose sum exceeds 1. Thus, since  $\mathfrak{s}$  is continuous and strictly proper, we can appeal to the main theorem of (Predd et al., 2009, Theorem 1), which says that, for any non-probabilistic credence function on an opinion set, there is another credence function on the same opinion set such that the sum of the local epistemic utilities of the latter exceeds the sum of the local epistemic utilities of the former at all worlds. Thus, there is a credence function  $c^*$  also defined on  $\mathcal{F}'$  such that

$$\sum_{X' \in \mathcal{F}'} \mathfrak{s}_{X'}(w(X'), c'(X')) < \sum_{X' \in \mathcal{F}'} \mathfrak{s}_{X'}(w(X'), c^*(X'))$$

for all  $w$ . Putting the last two inequalities together, we get:

$$\sum_{X \in \mathcal{F}} \mathfrak{s}_X(w(X), c(X)) < \sum_{X' \in \mathcal{F}'} \mathfrak{s}_{X'}(w(X'), c^*(X')) \quad (2)$$



And thus, by Total Epistemic Utilitarianism,  $c \prec_w c^*$  for all worlds  $w$ . So  $c$  is strictly dominated. Now, this holds for any credence function  $c$ ; *a fortiori* there is a credence  $x$  and an open proposition  $X$  such that any credence function that assigns  $x$  to  $X$  is strictly dominated.

This completes our proof. We showed that (i) entails (b); and (ii) together with the denial of (a) entails (b) as well.  $\square$

**Theorem 3** Suppose that  $\mathfrak{s}$  is truth-directed. And suppose that the epistemic utility ordering  $\preceq_w$  is governed by Average Epistemic Utilitarianism. Then, any credence function that assigns any non-extremal credences is strictly dominated.

*Proof.* Suppose  $c$  is defined on  $\mathcal{F}$  and  $c$  assigns some non-extremal credences. Then we will define a credence function  $c^*$  on an opinion set  $\mathcal{F}^*$  as follows. Let  $\top$  be a tautology. Suppose  $\mathfrak{s}(1,1) \geq \mathfrak{s}(0,0)$ . Then let  $\mathcal{F}^* = \{\top\}$  and let  $c^*(\top) = 1$ . Then, by Truth-Directedness, we have that, for all  $0 \leq x \leq 1$  and all worlds  $w$ ,  $\mathfrak{s}_\top(w(\top), 1) > \mathfrak{s}_X(w(X), x)$ . So, since  $c$  assigns some non-extremal credences,  $\mathfrak{s}_\top(w(\top), 1) \geq \mathfrak{s}_X(w(X), c(X))$  for all  $X$  in  $\mathcal{F}$  and all  $w$ , and  $\mathfrak{s}_\top(w(\top), 1) > \mathfrak{s}_X(w(X), c(X))$  for some  $X$  in  $\mathcal{F}$  and all  $w$ . Then

$$\frac{1}{|\mathcal{F}^*|} \sum_{X \in \mathcal{F}^*} \mathfrak{s}_X(w(X), c^*(X)) = \mathfrak{s}_\top(w(\top), 1) = \frac{1}{|\mathcal{F}|} \sum_{X \in \mathcal{F}} \mathfrak{s}(w(\top), 1) > \frac{1}{|\mathcal{F}|} \sum_{X \in \mathcal{F}} \mathfrak{s}_X(w(X), c(X))$$

And similarly, if  $\mathfrak{s}_\perp(0,0) \geq \mathfrak{s}(1,1)$ , then let  $\perp$  be a contradiction and define  $c^*$  on  $\mathcal{F}^* = \{\perp\}$  with  $c^*(\perp) = 0$ .  $\square$

**Theorem 6** There is no epistemic utility ordering  $\preceq_w$  on mixed options that contains an epistemic utility cycle amongst the unmixed options and that satisfies Stochastic Dominance.

*Proof.* Suppose  $c_1, \dots, c_n$  is an epistemic utility cycle on the unmixed options. Then let  $c'_1 = c_2, c'_2 = c_3, c'_3 = c_4, \dots, c'_{n-1} = c_n, c'_n = c_1$ . Then, since  $c_1, \dots, c_n$  is a cycle,

- (i)  $c_i \succeq_w c'_i$  for all  $i = 1, \dots, n$ ;
- (ii)  $c_i \succ_w c'_i$  for some  $i = 1, \dots, n$ ;

So by Stochastic Dominance, we have

$$\left\{ \frac{1}{n}, c_1; \dots; \frac{1}{n}, c_n \right\} \succ_w \left\{ \frac{1}{n}, c'_1; \dots; \frac{1}{n}, c'_n \right\} = \left\{ \frac{1}{n}, c_2; \dots; \frac{1}{n}, c_n; \frac{1}{n}, c_1 \right\} = \left\{ \frac{1}{n}, c_1; \dots; \frac{1}{n}, c_n \right\}$$

But this gives a contradiction, since there can be no option strictly better than itself.  $\square$

**Theorem 7** Benign Addition + Malign Addition + Transitivity + Partial Additivity<sup>+</sup> + Order Continuity + Utility Continuity + Truth-Directedness  $\Rightarrow$  Total Epistemic Utilitarianism

*Proof.* First, for a credence function  $c$  defined on  $\mathcal{F}$ , let  $\text{Total}_w(c) = \sum_{X \in \mathcal{F}} \mathfrak{s}_X(w(X), c(X))$ . Suppose  $c, c'$  are defined on  $\mathcal{F}, \mathcal{F}'$ , respectively. There are then four cases to consider:

- Suppose  $|\mathcal{F}| = |\mathcal{F}'|$ . And suppose  $\text{Total}_w(c) < / = / > \text{Total}_w(c')$ . Then, by Partial Additivity,  $c \prec_w / \sim_w / \succ_w c'$ .

- Suppose  $|\mathcal{F}| < |\mathcal{F}'|$  and  $\text{Total}(c) < \text{Total}(c')$ . Then it is possible to find  $c^*$  with the following properties: (i)  $c^*$  is obtained by extending  $c$  to an opinion set the same size as  $\mathcal{F}'$ ; (ii)  $c^*$  constitutes a benign addition to  $c$ ; (iii)  $\text{Total}_w(c^*) < \text{Total}_w(c')$ . Thus, by (i), (iii), and Partial Additivity<sup>+</sup>, we have  $c^* \prec_w c'$ . By Benign Addition and (ii), we have  $c \preceq_w c^*$ . And by Transitivity, we have  $c \prec_w c'$ , as required.
- Suppose  $|\mathcal{F}| < |\mathcal{F}'|$  and  $\text{Total}(c) > \text{Total}(c')$ . Then it is possible to find  $c^*$  with the following properties: (i)  $c^*$  is obtained by extending  $c$  to an opinion set the same size as  $\mathcal{F}'$ ; (ii)  $c^*$  constitutes a malign addition to  $c$ ; (iii)  $\text{Total}_w(c^*) > \text{Total}_w(c')$ . Thus, by (i), (iii), and Partial Additivity<sup>+</sup>, we have  $c^* \succ_w c'$ . By Malign Addition and (ii), we have  $c \succeq_w c^*$ . And by Transitivity, we have  $c \succ_w c'$ , as required.
- Suppose  $|\mathcal{F}| < |\mathcal{F}'|$  and  $\text{Total}(c) = \text{Total}(c')$ . Then let  $c^*$  be a credence function that is obtained by extending  $c$  to an opinion set  $\mathcal{F}^*$  that is the same size as  $\mathcal{F}'$  such that, for each  $X$  not in  $\mathcal{F}$ , the epistemic utility of the credence that  $c^*$  assigns to  $X$  is 0 — such a credence exists by Truth-Directedness and Utility Continuity. Thus,  $\text{Total}_w(c) = \text{Total}_w(c^*) = \text{Total}_w(c')$ . So, by Partial Additivity<sup>+</sup>,  $c^* \sim_w c'$ . We now show that  $c \sim_w c^*$ . Then there is a sequence of credence functions,  $c_1, c_2, \dots$  with the following properties: (i) each  $c_n$  is obtained by extending  $c$  to  $\mathcal{F}^*$ ; (ii) each  $c_n$  constitutes a malign addition to  $c$ ; (iii)  $\lim_{n \rightarrow \infty} c_n = c^*$ . By (i) and (ii) and Malign Addition,  $c_n \preceq_w c$ . By (iii) and Continuity,  $c^* \preceq_w c$ . Similarly, there is a sequence of credence functions,  $c_1, c_2, \dots$  with the following properties: (i) each  $c_n$  is obtained by extending  $c$  to  $\mathcal{F}^*$ ; (ii) each  $c_n$  constitutes a benign addition to  $c$ ; (iii)  $\lim_{n \rightarrow \infty} c_n = c^*$ . By (i) and (ii) and Benign Addition,  $c_n \succeq_w c$ . By (iii) and Continuity,  $c^* \succeq_w c$ . Putting these together, we have  $c^* \sim_w c$ . Combining this with  $c^* \sim_w c'$ , we get  $c \sim_w c'$ , as required.

This completes our proof, since the four cases exhaust all the possibilities.  $\square$

## References

- Arrhenius, G. (2000). An Impossibility Theorem for Welfarist Axiology. *Economics and Philosophy*, 16, 247–266.
- Arrhenius, G. (2003). The Very Repugnant Conclusion. In K. Segerberg, & R. Sliwinski (Eds.) *Logic, Law, Morality: thirteen essays in practical philosophy in honour of Lennart Åqvist*, no. 51 in Uppsala Philosophical Studies, (pp. 167–180). Uppsala: Uppsala University.
- Broome, J. (2004). *Weighing Lives*. Oxford: Oxford University Press.
- Carr, J. (ta). Epistemic Expansions. *Res Philosophica*.
- Chang, R. (2002). The Possibility of Parity. *Ethics*, 112, 659–88.
- Chang, R. (2005). Parity, Interval Value, and Choice. *Ethics*, 115, 331–50.
- Dalton, H. (1920). The Measurement of the Inequality of Incomes. *Economics Journal*, 30, 348–61.
- Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Goldman, A. (2002). The Unity of the Epistemic Virtues. In *Pathways to Knowledge: Private and Public*. New York: Oxford University Press.

- Greaves, H. (2013). Epistemic Decision Theory. *Mind*, 122(488), 915–952.
- Greaves, H., & Wallace, D. (2006). Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, 115(459), 607–632.
- Gustafsson, J. E. (2010). A Money-Pump for Acyclic Intransitive Preferences. *Dialectica*, 64(2), 251–257.
- Huemer, M. (2008). In Defence of Repugnance. *Mind*, 117(468), 899–933.
- Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, 65(4), 575–603.
- Joyce, J. M. (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In F. Huber, & C. Schmidt-Petri (Eds.) *Degrees of Belief*. Springer.
- Joyce, J. M. (ta). The True Consequences of Epistemic Consequentialism. In K. Ahlstrom-Vij, & J. Dunn (Eds.) *Epistemic Consequentialism*. Oxford: Oxford University Press.
- Kennedy, C. (2001). Polar Opposition and the Ontology of ‘Degrees’. *Linguistics and Philosophy*, 24, 33–70.
- Konek, J., & Levinstein, B. A. (ms). The Foundations of Epistemic Decision Theory. Unpublished manuscript.
- Lewis, D. (1980). A Subjectivist’s Guide to Objective Chance. In R. C. Jeffrey (Ed.) *Studies in Inductive Logic and Probability*, vol. II. Berkeley: University of California Press.
- Mellor, D. H. (1971). *The Matter of Chance*. Cambridge: Cambridge University Press.
- Ng, Y.-K. (1989). What should we do about future generations? Impossibility of Parfit’s theory X. *Economics and Philosophy*, 5, 235–53.
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- Pettigrew, R. (2013). A New Epistemic Utility Argument for the Principal Principle. *Episteme*, 10(1), 19–35.
- Pettigrew, R. (2014). Accuracy, Risk, and the Principle of Indifference. *Philosophy and Phenomenological Research*.
- Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.
- Pettigrew, R. (ta). Making Things Right: the true consequences of decision theory in epistemology. In K. Ahlstrom-Vij, & J. Dunn (Eds.) *Epistemic Consequentialism*. Oxford: Oxford University Press.
- Pigou, A. C. (1912). *Wealth and Welfare*. London: Macmillan and Co.
- Predd, J., Seiringer, R., Lieb, E. H., Osherson, D., Poor, V., & Kulkarni, S. (2009). Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions of Information Theory*, 55(10), 4786–4792.
- Rachels, S. (1998). Counterexamples to the transitivity of better than. *Australasian Journal of Philosophy*, 76(1), 71–83.
- Temkin, L. S. (1987). Intransitivity and the Mere Addition Paradox. *Philosophy and Public Affairs*, 16(2), 138–187.
- Williamson, T. (1998). Conditionalizing on Knowledge. *British Journal for the Philosophy of Science*, 49, 89–121.